PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : **G10L 3/02, 9/14**	A1	(11) International Publication Number: **WO 99/03095** (43) International Publication Date: 21 January 1999 (21.01.99)

(21) International Application Number: PCT/IB98/00871

(22) International Filing Date: 5 June 1998 (05.06.98)

(30) Priority Data:
97202163.8 11 July 1997 (11.07.97) EP
(34) Countries for which the regional or international application was filed: NL et al.

(71) Applicant: KONINKLIJKE PHILIPS ELECTRONICS N.V. [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(71) Applicant *(for SE only)*: PHILIPS AB [SE/SE]; Kottbygatan 7, Kista, S-164 85 Stockholm (SE).

(72) Inventors: TAORI, Rakesh; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). SLUIJTER, Robert, Johannes; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). GERRITS, Andreas, Johannes; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

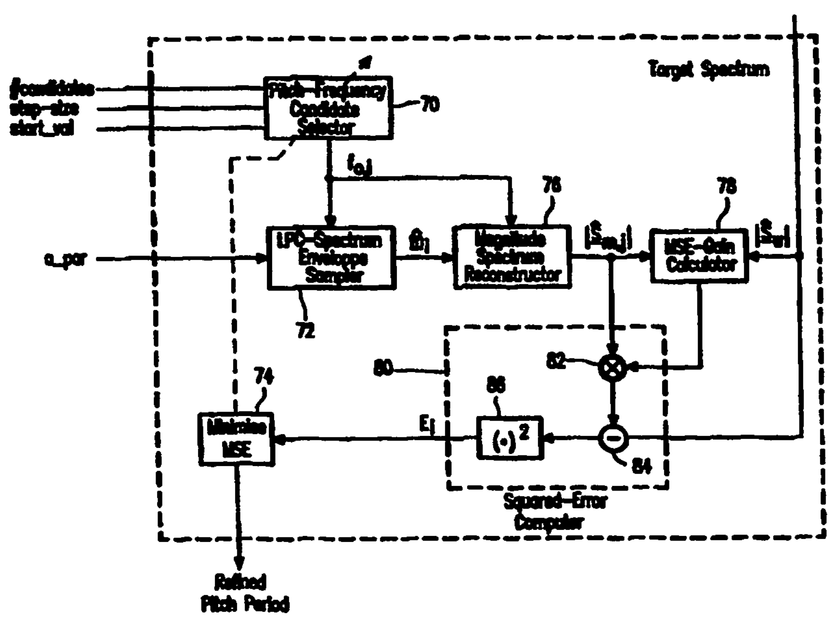
(74) Agent: DEGUELLE, Wilhelmus, H., G.; Internationaal Octrooibureau B.V., P.O. Box 220, NL-5600 AE Eindhoven (NL).

(81) Designated States: CN, JP, KR, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

PHN 17762 | DOSSIER

(54) Title: TRANSMITTER WITH AN IMPROVED HARMONIC SPEECH ENCODER



(57) Abstract

In a harmonic speech encoder (16) a speech signal to be encoded is represented by a plurality of LPC parameters which are determined by a LPC parameter computer (30), a pitch value and a gain value. The speech encoder comprises a (coarse) pitch estimator (38) for determining a coarse pitch, and a Refined Pitch Computer (32) to determine a Refined Pitch from the coarse pitch value. This determining of a refined pitch value is done in an analysis by synthesis way, in which a Refined Pitch value is selected which results in a minimum error measure between a representation of a synthetic speech signal and a representation of the original speech signal.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

Transmitter with an improved harmonic speech encoder.

The present invention is related to a transmitter with a speech encoder, said speech encoder comprises analysis means for determining a plurality of linear prediction coefficients from a speech signal, said analysis means comprises pitch determining means for determining a fundamental frequency of said speech signal, the analysis means further being arranged for determining an amplitude and a frequency of a plurality of harmonically related sinusoidal signals representing said speech signal from said plurality of linear prediction coefficients and said fundamental frequency.

The present invention is also related to a speech encoder, a speech encoding method and a tangible medium comprising a computer program implementing said method.

A transmitter according to the preamble is known from EP 259 950.

Such transmitters and speech encoders are used in applications in which speech signals have to be transmitted over a transmission medium with a limited transmission capacity or have to be stored on storage media with a limited storage capacity. Examples of such applications are the transmission of speech signals over the Internet, the transmission of speech signals from a mobile phone to a base station and vice versa and storage of speech signals on a CD-ROM, in a solid state memory or on a hard disk drive.

Different operating principles of speech encoders have been tried to achieve a reasonable speech quality at a modest bit rate. In one of these operating principles the speech signal is represented by a plurality of harmonically related sinusoidal signals. The transmitter comprises a speech encoder with analysis means for determining a pitch of the speech signal representing the fundamental frequency of said sinusoidal signals. The analysis means are also arranged for determining the amplitude of said plurality of sinusoidal signals.

The amplitudes of said plurality of sinusoidal signals can be obtained by determining prediction coefficients, calculating a frequency spectrum from said prediction coefficients, and sampling said frequency spectrum with the pitch frequency.

A problem with the known transmitters is that the quality of the reconstructed speech signal is lower than is expected.

An object of the present invention is to provide a transmitter according to the preamble which delivers an improved quality of the reconstructed speech.

Therefor the transmitter according to the invention is characterized in that the analysis means comprise pitch tuning means for tuning the fundamental frequency of said plurality of harmonically related signals in order to minimize a measure between a representation of said speech signal and a representation of said plurality of harmonically related sinusoidal signals, the transmitter comprising transmit means for transmitting a representation of said amplitudes and said fundamental frequency.

The present invention is based on the recognition that the combination of the amplitudes of the sinusoidal signals as determined by the analysis means and the pitch as determined by the pitch determining means do not constitute an optimal representation of the speech signal. By tuning the pitch in an analysis-by-synthesis like fashion it is possible to achieve an increased quality of the reconstructed speech signal without increasing the bit rate of the encoded speech signal.

The "analysis-by-synthesis" can be performed by comparing the original speech signal with a speech signal reconstructed on basis of the amplitudes and the actual pitch value. It is also possible to determine the spectrum of the original speech signal and to compare it with a spectrum determined from the amplitude of the sinusoidal signals and the pitch value.

An embodiment of the invention is characterized in that the determination of the amplitude and the frequency of a plurality of harmonically related speech signals is based on substantially unquantized prediction coefficients, in that the representation of said amplitudes comprises quantized prediction coefficients, and a gain factor which is determined on basis of the quantized prediction coefficients and said fundamental frequency.

From experiments it became clear that performing the "analysis by synthesis" on basis of the quantized prediction coefficients caused undesired artifacts in the reconstructed speech. Subsequently performed experiments have shown that, by using the unquantized prediction coefficients in the "analysis by synthesis" and calculating the gain factor from the quantised prediction coefficient and the (refined) fundamental frequency, these artifacts can be avoided.

An further embodiment of the invention is characterized in that the analysis means comprise initial pitch determining means for providing at least an initial pitch value for the pitch tuning means.

By using initial pitch determining means, it is possible to determine initial values for the analysis by synthesis lying close to the optimum pitch value. This will result in a decreased amount of computations required for finding said optimum pitch value.

The present invention will now be explained with reference to the drawing figures. Herein shows:

Fig. 1, a transmission system in which the present invention can be used;

Fig. 2, a speech encoder 4 according to the invention;

Fig. 3, a voiced speech encoder 16 according to the present invention;

Fig. 4, LPC computation means 30 for use in the voiced speech encoder 16 according to Fig. 3;

Fig. 5, pitch tuning means 32 for use in the speech encoder according to Fig. 3;

Fig. 6, an speech encoder 14 for unvoiced speech, for use in the speech encoder according to Fig. 2;

Fig. 7, a speech decoder 14 for use in the system according to Fig. 1;

Fig. 8, a voiced speech decoder 94 for use in the speech decoder 14;

Fig. 9, graphs of signals present at a number of points in the voiced speech decoder 94;

Fig. 10, an unvoiced speech decoder 96 for use in the speech decoder 14.

In the transmission system according to Fig. 1, a speech signal is applied to an input of a transmitter 2. In the transmitter 2, the speech signal is encoded in a speech encoder 4. The encoded speech signal at the output of the speech encoder 4 is passed to transmit means 6. The transmit means 6 are arranged for performing channel coding, interleaving and modulation of the coded speech signal.

The output signal of the transmit means 6 is passed to the output of the transmitter, and is conveyed to a receiver 5 via a transmission medium 8. At the receiver 5, the output signal of the channel is passed to receive means 7. These receive means 7 provide RF processing, such as tuning and demodulation, de-interleaving (if applicable)and channel decoding. The output signal of the receive means 7 is passed to the speech decoder 9 which converts its input signal to a reconstructed speech signal.

The input signal $s_s[n]$ of the speech encoder 4 according to Fig. 2, is filtered by a DC notch filter 10 to eliminate undesired DC offsets from the input. Said DC notch filter has a cut-off frequency (-3dB) of 15 Hz. The output signal of the DC notch filter 10 is applied to an

input of a buffer 11. The buffer 11 presents blocks of 400 DC filtered speech samples to a voiced speech encoder 16 according to the invention. Said block of 400 samples comprises 5 frames of 10 ms of speech (each 80 samples). It comprises the frame presently to be encoded, two preceding and two subsequent frames. The buffer 11 presents in each frame interval the most recently received frame of 80 samples to an input of a 200 Hz high pass filter 12. The output of the high pass filter 12 is connected to an input of a unvoiced speech encoder 14 and to an input of a voiced/unvoiced detector 28. The high pass filter 12 provides blocks of 360 samples to the voiced/unvoiced detector 28 and blocks of 160 samples (if the speech encoder 4 operates in a 5.2 kbit/sec mode) or 240 samples (if the speech encoder 4 operates in a 3.2 kbit/sec mode) to the unvoiced speech encoder 14. The relation between the different blocks of samples presented above and the output of the buffer 11 is presented in the table below.

Element	5.2 kbit/sec		3.2kbit/s	
	#samples	start	#samples	start
high pass filter 12	80	320	80	320
voiced/unvoiced detector 28	360	0 · · · 40	360	0 · · · 40
voiced speech encoder 16	400	0	400	0
unvoiced speech encoder 14	160	120	240	120
present frame to be encoded	80	160	80	160

The voiced/unvoiced detector 28 determines whether the current frame comprises voiced or unvoiced speech, and presents the result as a voiced/unvoiced flag. This flag is passed to a multiplexer 22, to the unvoiced speech encoder 14 and the voiced speech encoder 16. Dependent on the value of the voiced/unvoiced flag, the voiced speech encoder 16 or the unvoiced speech encoder 14 is activated.

In the voiced speech encoder 16 the input signal is represented as a plurality of harmonically related sinusoidal signals. The output of the voiced speech encoder provides a pitch value, a gain value and a representation of 16 prediction parameters. The pitch value and the gain value are applied to corresponding inputs of a multiplexer 22.

In the 5.2 kbit/sec mode the LPC computation is performed every 10 ms. In the 3.2 kbit/sec the LPC computation is performed every 20 ms, except when a transition

between unvoiced to voiced speech or vice versa takes place. If such a transition occurs, in the 3.2 kbit/sec mode the LPC calculation is also performed every 10 msec.

The LPC coefficients at the output of the voiced speech encoder are encoded by a Huffman encoder 24. The length of the Huffman encoded sequence is compared with the length of the corresponding input sequence by a comparator in the Huffman encoder 24. If the length of the Huffman encoded sequence is longer than the input sequence, it is decided to transmit the uncoded sequence. Otherwise it is decided to transmit the Huffman encoded sequence. Said decision is represented by a "Huffman bit" which is applied to a multiplexer 26 and to a multiplexer 22. The multiplexer 26 is arranged to pass the Huffman encoded sequence or the input sequence to the multiplexer 22 in dependence on the value of the "Huffman Bit". The use of the "Huffman bit" in combination with the multiplexer 26 has the advantage that it is ensured that the length of the representation of the prediction coefficients does not exceed a predetermined value. Without the use of the "Huffman bit" and the multiplexer 26 it could happen that the length of the Huffman encoded sequence exceeds the length of the input sequence in such an extent that the encoded sequence does not fit anymore in the transmit frame in which a limited number of bits are reserved for the transmission of the LPC coefficients.

In the unvoiced speech encoder 14 a gain value and 6 prediction coefficients are determined to represent the unvoiced speech signal. The 6 LPC coefficients are encoded by a Huffman encoder 18 which presents at its output a Huffman encoded sequence and a "Huffman bit". The Huffman encoded sequence and the input sequence of the Huffman encoder 18 are applied to a multiplexer 20 which is controlled by the "Huffman bit". The operation of the combination of the Huffman encoder 18 and the multiplexer 20 is the same as the operation of the Huffman encoder 24 and the multiplexer 20.

The output signal of the multiplexer 20 and the "Huffman bit" are applied to corresponding inputs of the multiplexer 22. The multiplexer 22 is arranged for selecting the encoded voiced speech signal or the encoded unvoiced speech signal, dependent on the decision of the voiced-unvoiced detector 28. At the output of the multiplexer 22 the encoded speech signal is available.

In the voiced speech encoder 16 according to Fig. 3, the analysis means according to the invention are constituted by the LPC Parameter Computer 30, the Refined Pitch Computer 32 and the Pitch Estimator 38. The speech signal s[n] is applied to an input of the LPC Parameter Computer 30. The LPC Parameter Computer 30 determines the prediction coefficients

a[i], the quantized prediction coefficients aq[i] obtained after quantizing, coding and decoding a[i], and LPC codes C[i], in which i can have values from 0-15.

The pitch determination means according to the inventive concept comprise initial pitch determining means, being here a pitch estimator 38, and pitch tuning means, being here a Pitch Range Computer 34 and a Refined Pitch Computer 32. The pitch estimator 38 determines a coarse pitch value which is used in the pitch range computer 34 for determining the pitch values which are to be tried in the pitch tuning means further to be referred to as Refined Pitch Computer 32 for determining the final pitch value. The pitch estimator 38 provides a coarse pitch period expressed in a number of samples. The pitch values to be used in the Refined Pitch Computer 32 are determined by the pitch range computer 34 from the coarse pitch period according to the table below.

Coarse pitch period p	Frequency (Hz)	Search Range	step-size	#candidates
$20 \le p \le 39$	400...200	p-3...p+3	0.25	24
$40 \le p \le 79$	200... 100	p-2...p+2	0.25	16
$80 \le p \le 200$	100...40	p	1	1

In the amplitude spectrum computer 36 a windowed speech signal S_{HAM} is determined from the signal s[i] according to:

$$S_{HAM}[i-120] = w_{HAM}[i] \cdot s[i] \qquad (1)$$

In (1) $w_{HAM}[i]$ is equal to:

$$w_{HAM} = 0.54 - 0.46\cos\left\{\frac{2\pi((i+0.5)-120)}{160}\right\} \quad ;120 \le i < 280 \qquad (2)$$

The windowed speech signal $s_{HAM}[i]$ is transformed to the frequency domain using a 512 point FFT. The spectrum S_w obtained by said transformation is equal to:

$$S_w[k] = \sum_{m=0}^{159} s_{HAM}[m] \cdot e^{-j2\pi km/512} \qquad (3)$$

BNSDOCID: <WO___9903095A1_I_>

The amplitude spectrum to be used in the Refined Pitch Computer 32 is calculated according to:

$$|S_w[k]| = \sqrt{(\Re\{S_w[k]\})^2 + (\Im\{S_w[k]\})^2} \tag{4}$$

The Refined Pitch Computer 32 determines from the a-parameters provided by the LPC Parameter Computer 30 and the coarse pitch value a refined pitch value which results in a minimum error signal between the amplitude spectrum according to (4) and the amplitude spectrum of a signal comprising a plurality of harmonically related sinusoidal signals of which the amplitudes have been determined by sampling the LPC spectrum by said refined pitch period.

In the gain computer 40 the optimum gain to match the target spectrum accurately is calculated from the spectrum of the re-synthesized speech signal using the quantized a- parameters, instead of using the non-quantized a-parameters as is done in the Refined Pitch Computer 32.

At the output of the voiced speech encoder 40 the 16 LPC codes, the refined pitch and the gain calculated by the Gain Computer 40 are available. The operation of the LPC parameter computer 30 and the Refined Pitch Computer 32 are explained below in more detail.

In the LPC computer 30 according to Fig. 4, a window operation is performed on the signal s[n] by a window processor 50. According to one aspect of the present invention, the analysis length is dependent on the value of the voiced/unvoiced flag. In the 5.2 kbit/sec mode, the LPC computation is performed every 10 msec. In the 3.2 kbit/sec mode, the LPC calculation is performed every 20 msec, except during transitions from voiced to unvoiced or vice versa. If such a transition is present, the LPC calculation is performed every 10 msec.

In the following table the number of samples involved with the determination of the prediction coefficients are given.

Bit Rate and Mode	Analysis length N_A and samples involved	Update interval
5.2 kbit/s	160 (120-280)	10 ms
3.2 kbit/s (transition)	160 (120-280)	10 ms
3.2 kbit/s (no transition)	240 (120-360)	20 ms

BNSDOCID: <WO___9903095A1_I_>

For the window in the 5.2 kbit/sec case and in the 3.2 kbit/s case where a transition is present, can be written:

$$w_{HAM} = 0.54 - 0.46\cos\left\{\frac{2\pi((i+0.5)-120)}{160}\right\} \quad ;120 \le i < 280 \qquad (5)$$

For the windowed speech signal is found:

$$s_{HAM}[i-120] = w_{HAM}[i]\cdot s[i] \quad ;120 \le i < 280 \qquad (6)$$

If in the 3.2 kbit/s case no transition is present, a flat top portion of 80 samples is introduced in the middle of the window thereby extending the window to span 240 samples starting at sample 120 and ending before sample 360. In this way a window w'_{HAM} is obtained according to:

$$w'_{HAM} = \begin{cases} w_{HAM}[i] & ;120 \le i < 200 \\ 1 & ;200 \le i < 280 \\ w_{HAM}[i] & ;280 \le i < 360 \end{cases} \qquad (7)$$

for the windowed speech signal the following can be written.

$$s_{HAM}[i-120] = w_{HAM}[i]\cdot s[i] \quad ;120 \le i < 360 \qquad (8)$$

The Autocorrelation Function Computer 58 determines the autocorrelation function R_{SS} of the windowed speech signal. The number of correlation coefficients to be calculated is equal to the number of prediction coefficients + 1. If a voiced speech frame is present, the number of autocorrelation coefficients to be calculated is 17. If an unvoiced speech frame is present, the number of autocorrelation coefficients to be calculated is 7. The presence of a voiced or unvoiced speech frame is signaled to the Autocorrelation Function Computer 58 by the voiced/unvoiced flag.

The autocorrelation coefficients are windowed with a so-called lag-window in order to obtain some spectral smoothing of the spectrum represented by said autocorrelation coefficients. The smoothed autocorrelation coefficients $\rho[i]$ are calculated according to :

$$\rho[i] = R_{SS}[i] \cdot \exp\left(\frac{-\pi f_\mu i}{8000}\right) \quad ; 0 \le i \le P \tag{9}$$

In **(9)** f_μ is the spectral smoothing constant having a value of 46.4 Hz. The windowed autocorrelation values $\rho[i]$ are passed to the Schur recursion module 62 which calculates the reflection coefficients $k[1]$ to $k[P]$ in a recursive way. The Schur recursion is well known to those skilled in the art.

In a converter 66 the P reflection coefficients $\rho[i]$ are transformed into a-parameters for use in the Refined Pitch Computer 32 in Fig. 3. In a quantizer 64 the reflection coefficients are converted into Log Area Ratios, and these Log Area Ratios are subsequently uniformly quantized. The resulting LPC codes $C[1] \cdots\cdots C[P]$ are passed to the output of the LPC parameter computer for further transmission.

In the local decoder 54 the LPC codes $C[1] \cdots\cdots C[P]$ are converted into reconstructed reflection coefficients $\hat{k}[i]$ by a reflection coefficient reconstructor 54. Subsequently the reconstructed reflection coefficients $\hat{k}[i]$ are converted into (quantized) a-parameters by the Reflection Coefficient to a-parameter converter 56.

This local decoding is performed in order to have the same a-parameters available in the speech encoder 4 and the speech decoder 14.

In the Refined Pitch Computer 32 according to Fig. 5, a Pitch Frequency Candidate Selector 70 determines from the number of candidates, the start value and the step size as received from the Pitch Range Computer 34 the candidate pitch values to be used in the Refined Pitch Computer 32. For each of the candidates, the Pitch Frequency Candidate Selector 70 determines a fundamental frequency $f_{0,i}$.

Using the candidate frequency $f_{0,i}$ the spectral envelope described by the LPC coefficients is sampled at harmonic locations by the Spectrum Envelope Sampler 72. For $m_{i,k}$ being the amplitude of the k^{th} harmonic of the i^{th} candidate $f_{0,i}$ can be written:

$$m_{i,k} = \left|\frac{1}{A(z)}\right|_{z=2\pi k \cdot f_{0,i}} \tag{10}$$

In **(10)**, A(z) is equal to :

$$A(z)=1+a_1\cdot z^{-1}+a_2\cdot z^{-2}+\cdots+a_P\cdot z^{-P} \quad (11)$$

With $z=e^{j\theta_{i,k}}=\cos\theta_{i,k}+j\cdot\sin\theta_{i,k}$ and $\theta_{i,k}=2\pi kf_{0,i}$ **(11)** changes into:

$$A(z)\big|_{\theta=\theta_{i,k}}=1+a_1(\cos\theta_{i,k}+j\cdot\sin\theta_{i,k})+\cdots+a_P(\cos\theta_{P,k}+j\cdot\sin\theta_{P,k}) \quad (12)$$

By splitting **(12)** into real and imaginary parts, the amplitudes $m_{i,k}$ can be obtained according to:

$$m_{i,k}=\frac{1}{\sqrt{R^2(\theta_{i,k})+I^2(\theta_{i,k})}} \quad (13)$$

where

$$R(\theta_{i,k})=1+a_1(\cos\theta_{i,k})+\cdots+a_P(\cos\theta_{i,k}) \quad (14)$$

and

$$I(\theta_{i,k})=1+a_1(\sin\theta_{i,k})+\cdots+a_P(\sin\theta_{i,k}) \quad (15)$$

The candidate spectrum $|\hat{S}_{w,i}|$ is determined by convolving the spectral lines $m_{i,k}$ ($1\le k\le L$) with a spectral window function W which is the 8192 point FFT of the 160 points Hamming window according to **(5)** or **(7)**, dependent on the current operating mode of the encoder. It is observed that the 8192 points FFT can be pre-calculated and that the result can be stored in ROM. In the convolving process a downsampling operation is performed because the candidate spectrum has to be compared with 256 points of the reference spectrum, making calculation of more than 256 points useless. Consequently for $|\hat{S}_{w,i}|$ can be written:

$$|\hat{S}_{w,i}[f]|=\sum_{k=1}^{L}m_{i,k}\cdot W(16\cdot f-k\cdot f_{0,i}) \quad ;0\le f<256 \quad (16)$$

Expression **(16)** gives only the general shape of the amplitude spectrum for pitch candidate i, but not its amplitude. Consequently the spectrum $|\hat{S}_{w,i}|$ has to be corrected by a gain factor g_i which is calculated by a MSE-gain Calculator 78 according to:

$$g_i = \frac{\sum_{j=0}^{256} S_w[j] \cdot \hat{S}_{w,i}[j]}{\sum_{j=0}^{256} (S_w[j])^2} \qquad (17)$$

A multiplier 82 is arranged for scaling the spectrum $|\hat{S}_{w,i}|$ with the gain factor g_i. A subtracter 84 computes the difference between the coefficients of the target spectrum as determined by the Amplitude Spectrum Computer 36 and the output signal of the multiplier 82. Subsequently a summing squarer computes a squared error signal E_i according to:

$$E_i = E(f_{0,i}) = \sum_{j=0}^{255} \left(|S_w[j]| - g_i \cdot |\hat{S}_{w,i}[j]| \right)^2 \qquad (18)$$

The candidate fundamental frequency, $f_{0,i}$ that results in the minimum value is selected as the refined fundamental frequency or refined pitch. In the encoder according to the present example, a total of 368 pitch periods are possible requiring 9 bits for encoding. The pitch is updated every 10 msec independent of the mode of the speech encoder. In the gain calculator 40 according to Fig. 3, the gain to be transmitted to the decoder is calculated in the same way as is described above with respect to the gain g_i , but now the quantized a-parameters are used instead of the unquantized a-parameters which are used when calculating the gain g_i. The gain factor to be transmitted to the decoder is non-linearly quantized in 6 bits, such that for small values of g_i small quantization steps are used, and for larger values of g_i larger quantization steps are used.

In the unvoiced speech encoder 14 according to Fig. 6, the operation of the LPC parameter computer 82 is similar to the operation of the LPC parameter computer 30 according to Fig. 4. The LPC parameter computer 82 operates on the high pass filtered speech signal instead of on the original speech signal as in done by the LPC parameter computer 30. Further the prediction order of the LPC computer 82 is 6 instead of 16 as is used in the LPC parameter pitch computer 30.

The time domain window processor 84 calculates a Hanning windowed speech signal according to:

$$s_W[n] = s[n] \cdot \left(0.5 - 0.5\cos\left(\frac{2\cdot\pi(i+0.5)-120)}{160}\right)\right) \quad ;120 \le i < 280 \tag{19}$$

In an RMS value computer 86 an average value g_{UV} of the amplitude of a speech frame is calculated according to:

$$g_{uv} = \frac{1}{4}\sqrt{\frac{1}{N}\sum_{i=0}^{159} s_w^2[i]} \tag{20}$$

The gain factor g_{uv} to be transmitted to the decoder is non-linearly quantized in 5 bits, such that for small values of g_{uv} small quantization steps are used, and for larger values of g_{uv} larger quantization steps are used. No excitation parameters are determined by the unvoiced speech encoder 14.

In the speech decoder 14 according to Fig. 7, the Huffman encoded LPC codes and a voiced/unvoiced flag are applied to a Huffman decoder 90. The Huffman decoder 90 is arranged for decoding the Huffman encoded LPC codes according to the Huffman table used by the Huffman encoder 18 if the voiced/unvoiced flag indicates an unvoiced signal. The Huffman decoder 90 is arranged for decoding the Huffman encoded LPC codes according to the Huffman table used by the Huffman encoder 24 if the voiced/unvoiced flag indicates a voiced signal. In dependence on the value of the Huffman bit, the received LPC codes are decoded by the Huffman decoder 90 or passed directly to a demultiplexer 92. The gain value and the received refined pitch value are also passed to the demultiplexer 92.

If the voiced/unvoiced flag indicates a voiced speech frame, the refined pitch, the gain and the 16 LPC codes are passed to a harmonic speech synthesizer 94. If the voiced/unvoiced flag indicates an unvoiced speech frame, the gain and the 6 LPC codes are passed to an unvoiced speech synthesizer 96. The synthesized voiced speech signal $\hat{s}_{v,k}[n]$ at the output of the harmonic speech synthesizer 94 and the synthesized unvoiced speech signal $\hat{s}_{uv,k}[n]$ at the output of the unvoiced speech synthesizer 96 are applied to corresponding inputs of a multiplexer 98.

In the voiced mode, the multiplexer 98 passes the output signal $\hat{s}_{v,k}[n]$ of the Harmonic Speech Synthesizer 94 to the input of the Overlap and Add Synthesis block 100. In the unvoiced mode, the multiplexer 98 passes the output signal $\hat{s}_{uv,k}[n]$ of the Unvoiced Speech

Synthesizer 96 to the input of the Overlap and Add Synthesis block 100. In the Overlap and Add Synthesis block 100, partly overlapping voiced and unvoiced speech segments are added. For the output signal $\hat{s}[n]$ of the Overlap and Add Synthesis Block 100 can be written:

$$\hat{s}[n] = \begin{cases} \hat{s}_{uv,k-1}[n+N_s/2] + \hat{s}_{uv,k}[n] & ; v_{k-1}=0, v_k=0 \\ \hat{s}_{uv,k-1}[n+N_s/2] + \hat{s}_{v,k}[n] & ; v_{k-1}=0, v_k=1 \\ \hat{s}_{v,k-1}[n+N_s/2] + \hat{s}_{uv,k}[n] & ; v_{k-1}=1, v_k=0 \\ \hat{s}_{v,k-1}[n+N_s/2] + \hat{s}_{v,k}[n] & ; v_{k-1}=1, v_k=1 \end{cases}$$

In **(21)** N_S is the length of the speech frame, $v_{k-1}$is the voiced/unvoiced flag for the previous speech frame, and v_k is the voiced/unvoiced flag for the current speech frame.

The output signal $\hat{s}[n]$ of the Overlap and Block is applied to a postfilter 102. The postfilter is arranged for enhancing the perceived speech quality by suppressing noise outside the formant regions.

In the voiced speech decoder 94 according to Fig. 8, the encoded pitch received from the demultiplexer 92 is decoded and converted into a pitch period by a pitch decoder 104. The pitch period determined by the pitch decoder 104 is applied to an input of a phase synthesizer 106, to an input of a Harmonic Oscillator Bank 108 and to a first input of a LPC Spectrum Envelope Sampler 110.

The LPC coefficients received from the demultiplexer 92 is decoded by the LPC decoder 112. The way of decoding the LPC coefficients depends on whether the current speech frame contains voiced or unvoiced speech. Therefore the voiced/unvoiced flag is applied to a second input of the LPC decoder 112. The LPC decoder passes the quantized a-parameters to a second input of the LPC Spectrum envelope sampler 110. The operation of the LPC Spectral Envelope Sampler 112 is described by **(13)**, **(14)** and **(15)** because the same operation is performed in the Refined Pitch Computer 32.

The phase synthesizer 106 is arranged to calculate the phase $\varphi_k[i]$of the i^{th} sinusoidal signal of the L signals representing the speech signal. The phase $\varphi_k[i]$ is chosen such that the i^{th} sinusoidal signal remains continuous from one frame to a next frame. The voiced speech signal is synthesized by combining overlapping frames, each comprising 160 windowed samples. There is a 50% overlap between two adjacent frames as can be seen from graph 118 and

graph 122 in Fig. 9 . In graphs 118 and 122 the used window is shown in dashed lines. The phase synthesizer is now arranged to provide a continuous phase at the position where the overlap has its largest impact. With the window function used here this position is at sample 119. For the phase $\varphi_k[i]$ of the current frame can now be written:

$$\varphi_k[i] = \varphi_{k-1}[i] + i \cdot 2\pi \cdot f_{0,k-1} \frac{3N_s}{4} - i \cdot 2\pi \cdot f_{0,k} \frac{N_s}{4} \quad ; 1 \le i \le 100 \tag{22}$$

In the currently described speech encoder the value of N_s is equal to 160. For the very first voiced speech frame, the value of $\varphi_k[i]$ is initialized to a predetermined value. The phases $\varphi_k[i]$ are always updated, even if an unvoiced speech frame is received. In said case,

$f_{0,k}$ is set to 50 Hz.

The harmonic oscillator bank 108 generates the plurality of harmonically related signals $\hat{s}'_{v,k}[n]$ that represents the speech signal. This calculation is performed using the harmonic amplitudes $\hat{m}[i]$, the frequency $\hat{f}_0$ and the synthesized phases $\hat{\varphi}[i]$ according to:

$$\hat{s}'_{v,k}[n] = \sum_{i=1}^{L} \hat{m}[i] \cos\{(i \cdot 2\pi \cdot f_0) \cdot n + \hat{\varphi}[i]\} \quad ; 0 \le n < N_s \tag{23}$$

The signal $\hat{s}'_{v,k}[n]$ is windowed using a Hanning window in the Time Domain Windowing block 114. This windowed signal is shown in graph 120 of Fig. 9. The signal $\hat{s}'_{v,k+1}[n]$ is windowed using a Hanning window being $N_s/2$ samples shifted in time. This windowed signal is shown in graph 124 of Fig. 9. The output signals of the Time Domain Windowing Block 144 is obtained by adding the above mentioned windowed signals. This output signal is shown in graph 126 of Fig. 9. A gain decoder 118 derives a gain value g_v from its input signal, and the output signal of the Time Domain Windowing Block 114 is scaled by said gain factor g_v by the Signal Scaling Block 116 in order to obtain the reconstructed voiced speech signal $\hat{s}_{v,k}$.

In the unvoiced speech synthesizer 96, the LPC codes and the voiced/unvoiced flag are applied to an LPC Decoder 130. The LPC decoder 130 provides a plurality of 6 a-parameters to an LPC Synthesis filter 134. An output of a Gaussian White-Noise Generator 132 is connected to an input of the LPC synthesis filter 143. The output signal of the LPC synthesis filter 134 is windowed by a Hanning window in the Time Domain Windowing Block 140.

An Unvoiced Gain Decoder 136 derives a gain value $\hat{g}_{uv}$ representing the desired energy of the present unvoiced frame. From this gain and the energy of the windowed signal, a scaling factor $\hat{g}'_{uv}$ for the windowed speech signal gain is determined in order to obtain a speech signal with the correct energy. For this scaling factor can be written:

$$\hat{g}'_{uv} = \sqrt{\frac{\hat{g}_{uv}}{\sum_{n=0}^{N_s-1} (\hat{s}'_{uv,k}[n] \cdot w[n])^2}} \quad (24)$$

The Signal Scaling Block 142 determines the output signal $\hat{s}_{uv,k}$ by multiplying the output signal of the time domain window block 140 by the scaling factor $\hat{g}'_{uv}$.

The presently described speech encoding system can be modified to require a lower bitrate or a higher speech quality. An example of a speech encoding system requiring a lower bitrate is a 2kbit/sec encoding system. Such a system can be obtained by reducing the number of prediction coefficients used for voiced speech from 16 to 12, and by using differential encoding of the prediction coefficients, the gain and the refined pitch. Differential coding means that the date to be encoded is not encoded individually, but that only the difference between corresponding data from subsequent frames is transmitted. At a transition from voiced to unvoiced speech or vice versa, in the first new frame all coefficients are encoded individually in order to provide a starting value for the decoding.

It is also possible to obtain a speech coder with an increased speech quality at a bit rate of 6kbit/s. The modifications are here the determination of the phase of the first 8 harmonics of the plurality of harmonically related sinusoidal signals. The phase $\varphi[i]$ is calculated according to:

$$\varphi[i] = \arctan \frac{I(\theta_i)}{R(\theta_i)} \quad (25)$$

Herein is $\theta_i = 2\pi f_0 \cdot i$. $R(\theta_i)$ en $I(\theta_i)$ are equal to:

$$R(\theta_i) = \sum_{n=0}^{N-1} s_w[n] \cdot \cos(\theta_i \cdot n) \quad (26)$$

and

$$I(\theta_i) = -\sum_{n=0}^{N-1} s_w[n] \cdot \sin(\theta_i \cdot n) \qquad (27)$$

The 8 phases $\varphi[i]$ obtained so are uniformly quantised to 6 bits and included in the output bitstream.

A further modification in the 6 kbit/sec encoder is the transmission of additional gain values in the unvoiced mode. Normally every 2 msec a gain is transmitted instead of once per frame. In the first frame directly after a transition, 10 gain values are transmitted, 5 of them representing the current unvoiced frame, and 5 of them representing the previous voiced frame that is processed by the unvoiced speech encoder. The gains are determined from 4 msec overlapping windows.

It is observed that the number of LPC coefficients is 12 and that where possible differential encoding is utilized.

Claims

1. Transmitter with a speech encoder, said speech encoder comprises analysis means for determining a plurality of linear prediction coefficients from a speech signal, said analysis means comprises pitch determining means for determining a fundamental frequency of said speech signal, the analysis means further being arranged for determining an amplitude and a frequency of a plurality of harmonically related sinusoidal signals representing said speech signal from said plurality of linear prediction coefficients and said fundamental frequency, characterized in that the analysis means comprise pitch tuning means for tuning the fundamental frequency of said plurality of harmonically related signals in order to minimize a measure between a representation of said speech signal and a representation of said plurality of harmonically related sinusoidal signals, the transmitter comprising transmit means for transmitting a representation of said amplitudes and said fundamental frequency.

2. Transmission system according to claim 1, characterized in that the determination of the amplitude and the frequency of a plurality of harmonically related speech signals is based on substantially unquantized prediction coefficients, in that the representation of said amplitudes comprises quantized prediction coefficients, and a gain factor which is determined on basis of the quantized prediction coefficients and said fundamental frequency.

3. Transmitter according to claim 1 or 2, characterized in that the analysis means comprise initial pitch determining means for providing at least an initial pitch value for the pitch tuning means.

4. Transmitter according to one of the previous claims, characterized in that the speech encoder comprises spectrum analysis means for determining a frequency spectrum of the speech signal, and in that the pitch tuning means are arranged to minimize a difference between a spectrum derived from said amplitudes and fundamental frequency and the spectrum of the frequency spectrum of the speech signal.

BNSDOCID: <WO___9903095A1_I_>

5. Speech encoder comprising analysis means for determining a plurality of linear prediction coefficients from a speech signal, said analysis means comprises pitch determining means for determining a fundamental frequency of said speech signal, the analysis means further being arranged for determining an amplitude and a frequency of a plurality of harmonically related sinusoidal signals representing said speech signal from said plurality of linear prediction coefficients and said fundamental frequency, characterized in that the analysis means comprise pitch tuning means for tuning the fundamental frequency of said plurality of harmonically related signals in order to minimize an difference measure between a representation of said speech signal and a representation of said plurality of harmonically related sinusoidal signals, the transmitter comprising transmit means for transmitting a representation of said amplitudes and said fundamental frequency.

6. Speech encoder according to claim 5, characterized in that the analysis means comprise initial pitch determining means for providing at least an initial pitch value for the pitch tuning means.

7. Speech encoder according to claim 5 or 6, characterized in that the speech encoder comprises spectrum analysis means for determining a frequency spectrum of the speech signal, and in that the pitch tuning means are arranged to minimize a difference between a spectrum derived from said amplitudes and fundamental frequency and the spectrum of the frequency spectrum of the speech signal.

8. Speech encoding method comprising determining a plurality of linear prediction coefficients from a speech signal, determining a fundamental frequency of said speech signal, determining an amplitude and a frequency of a plurality of harmonically related sinusoidal signals representing said speech signal from said plurality of linear prediction coefficients and said fundamental frequency, characterized in that the method comprises tuning the fundamental frequency of said plurality of harmonically related signals in order to minimize an difference measure between a representation of said speech signal and a representation of said plurality of harmonically related sinusoidal signals.

BNSDOCID: <WO___9903095A1_I_>

9. Method according to claim 8, characterized in that the method comprises providing at least an initial pitch value for the pitch tuning means.

10. Method according to claim 8 or 9, characterized in that the method comprises determining a frequency spectrum of the speech signal, and in that the method comprises minimizing a difference between a spectrum derived from said amplitudes and fundamental frequency and the spectrum of the frequency spectrum of the speech signal.

11. Tangible medium comprising a computer program for executing a speech encoding method comprising, determining a plurality of linear prediction coefficients from a speech signal, determining a fundamental frequency of said speech signal, determining an amplitude and a frequency of a plurality of harmonically related sinusoidal signals representing said speech signal from said plurality of linear prediction coefficients and said fundamental frequency, characterized in that the method comprises tuning the fundamental frequency of said plurality of harmonically related signals in order to minimize an difference measure between a representation of said speech signal and a representation of said plurality of harmonically related sinusoidal signals.

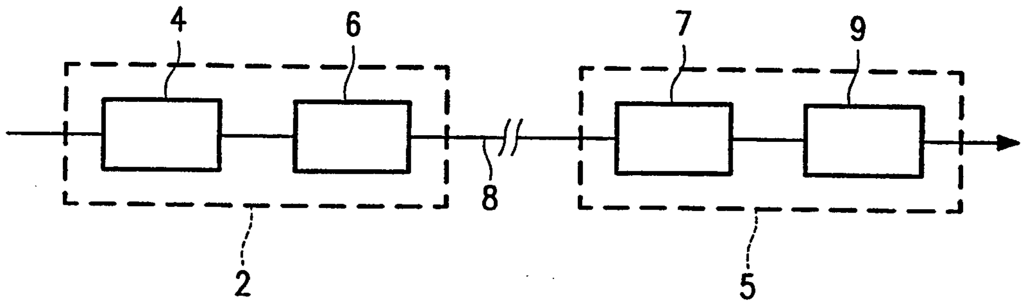


FIG. 1

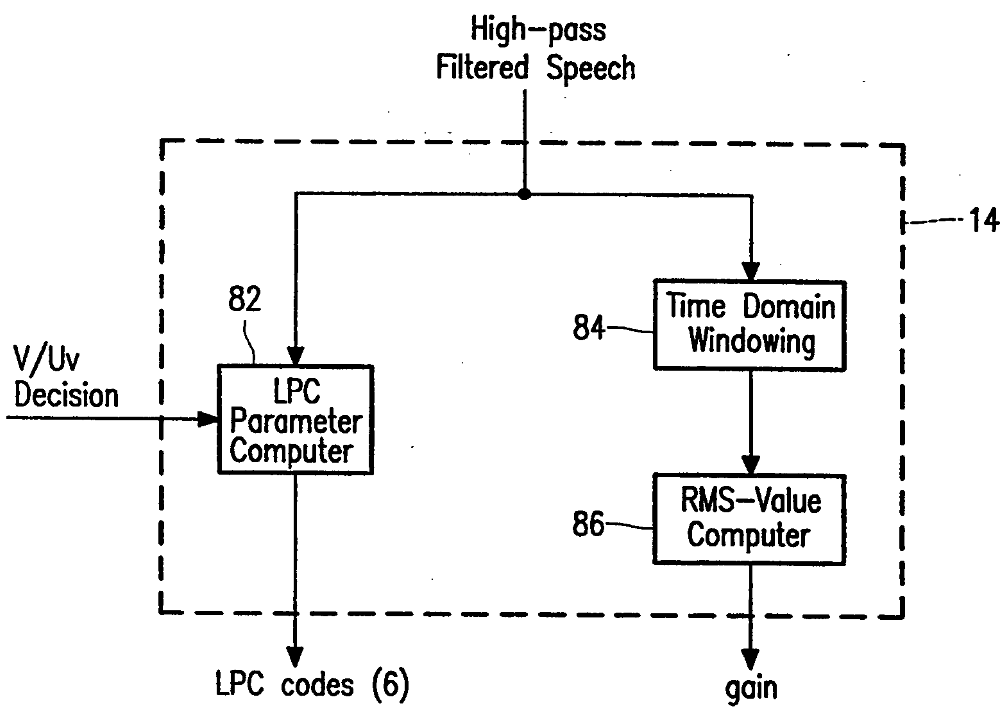


FIG. 6

BNSDOCID: <WO___9903095A1_I_>

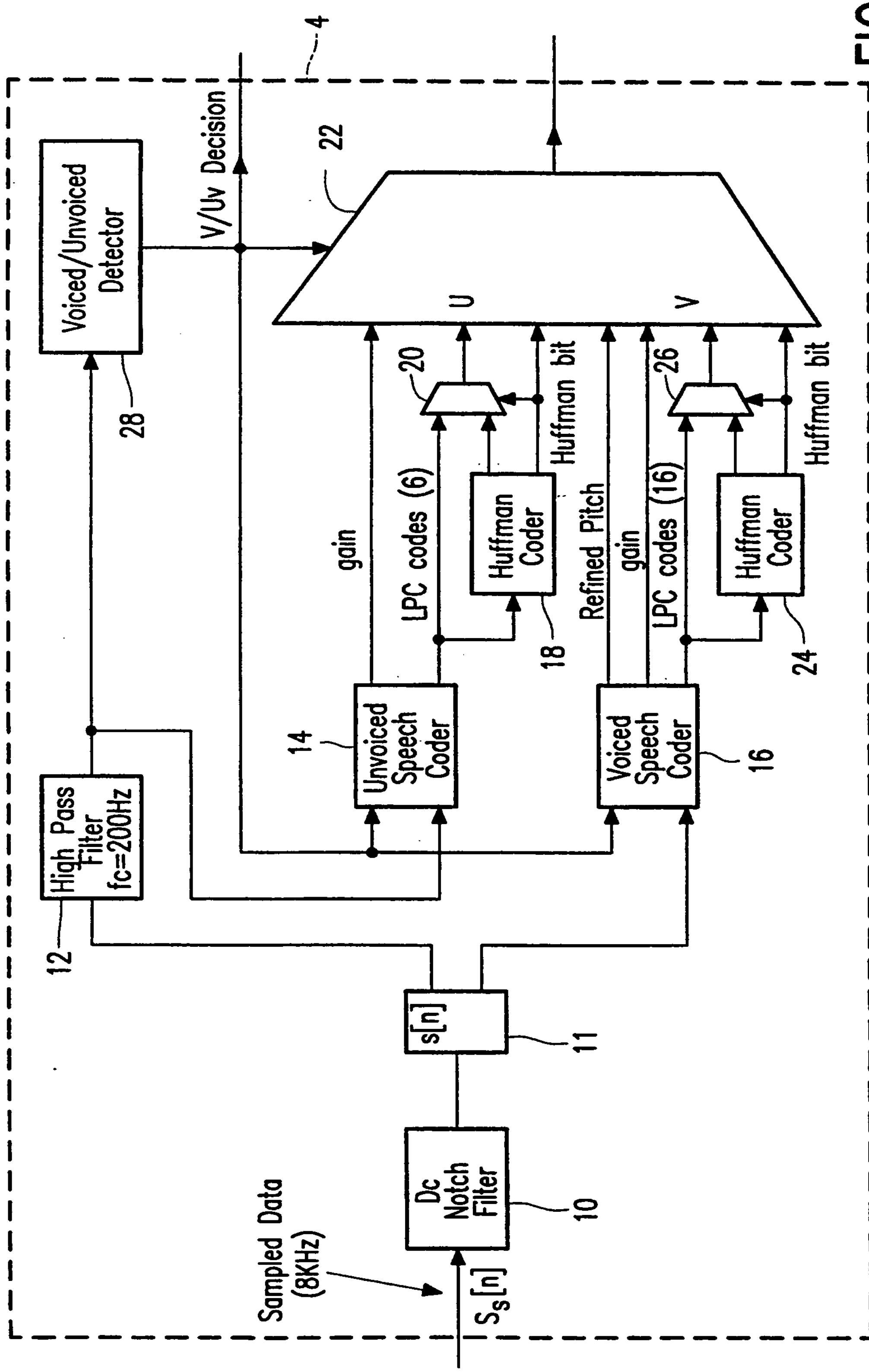

4
Voiced/Unvoiced Detector
28
V/Uv Decision
22
U
V
High Pass Filter fc=200Hz
12
Unvoiced Speech Coder
14
gain
LPC codes (6)
Huffman Coder
18
20
Huffman bit
Voiced Speech Coder
16
Refined Pitch
gain
LPC codes (16)
Huffman Coder
24
26
Huffman bit
s[n]
11
Dc Notch Filter
10
Sampled Data (8KHz)
$S_s[n]$

FIG. 2

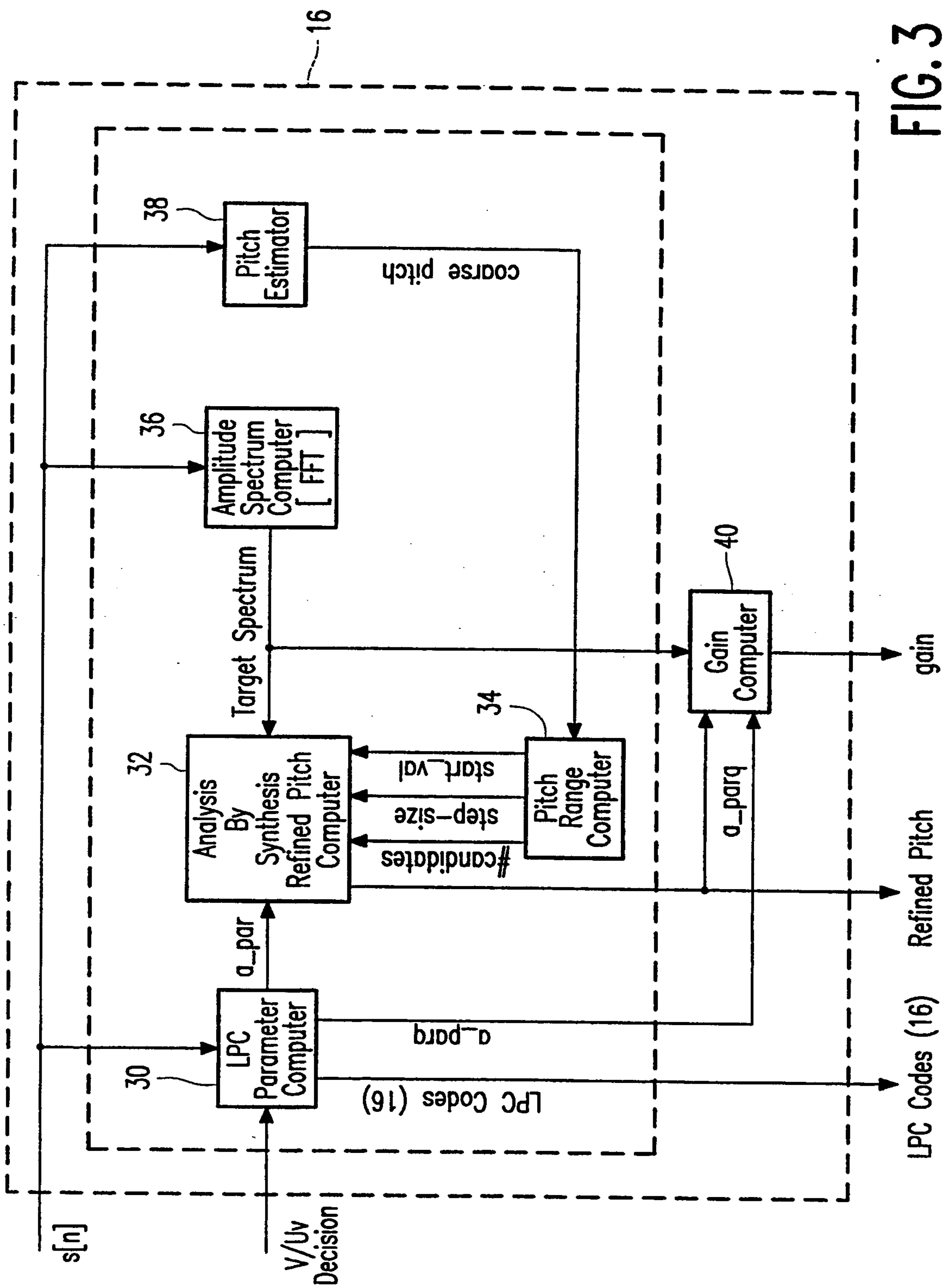

16
38
Pitch Estimator
coarse pitch
36
Amplitude Spectrum Computer [FFT]
Target Spectrum
40
Gain Computer
gain
34
Pitch Range Computer
start_val
step-size
#candidates
32
Analysis By Synthesis Refined Pitch Computer
a_par
a_parq
Refined Pitch
30
LPC Parameter Computer
a_parq
LPC Codes (16)
LPC Codes (16)
s[n]
V/Uv Decision

FIG. 3

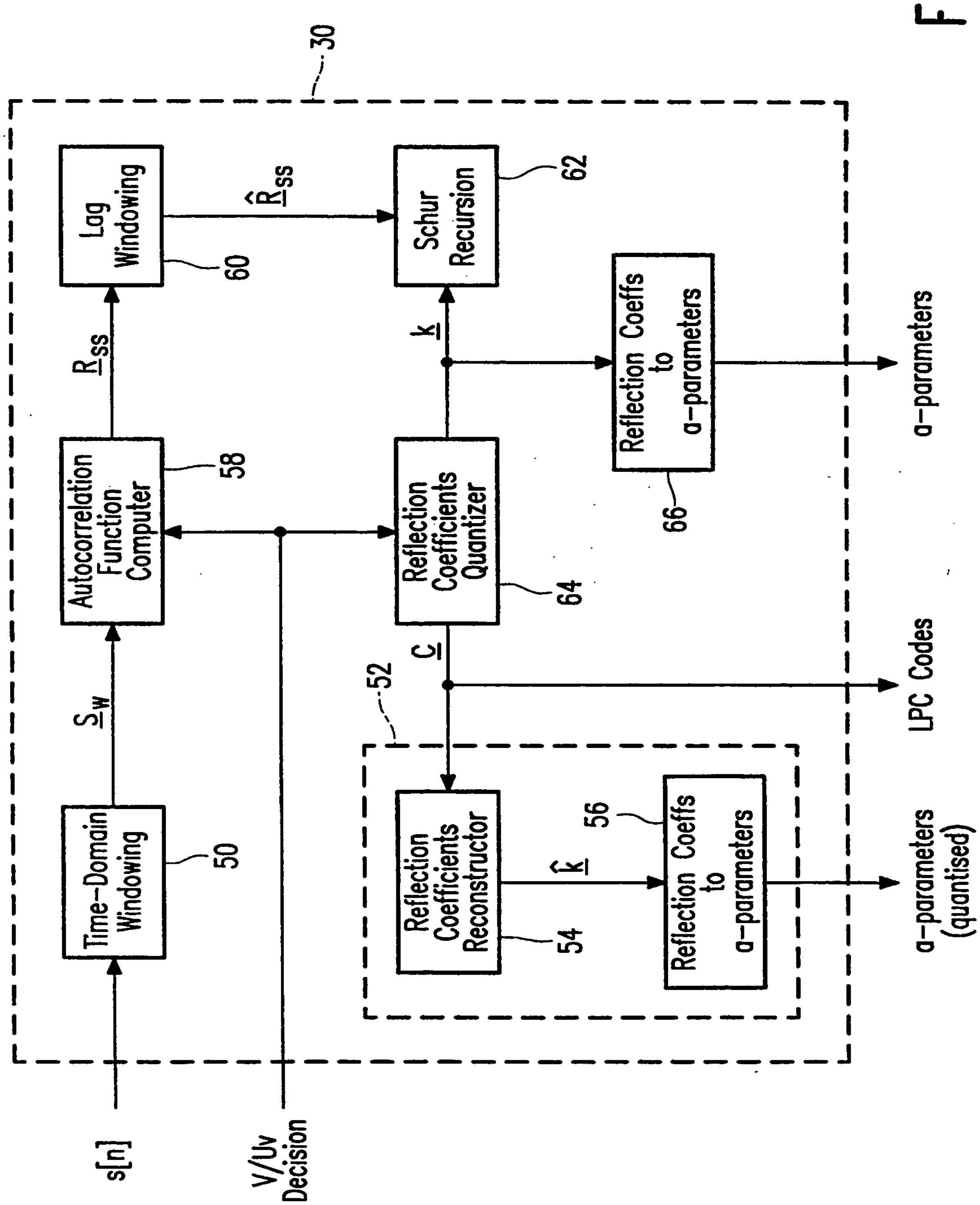
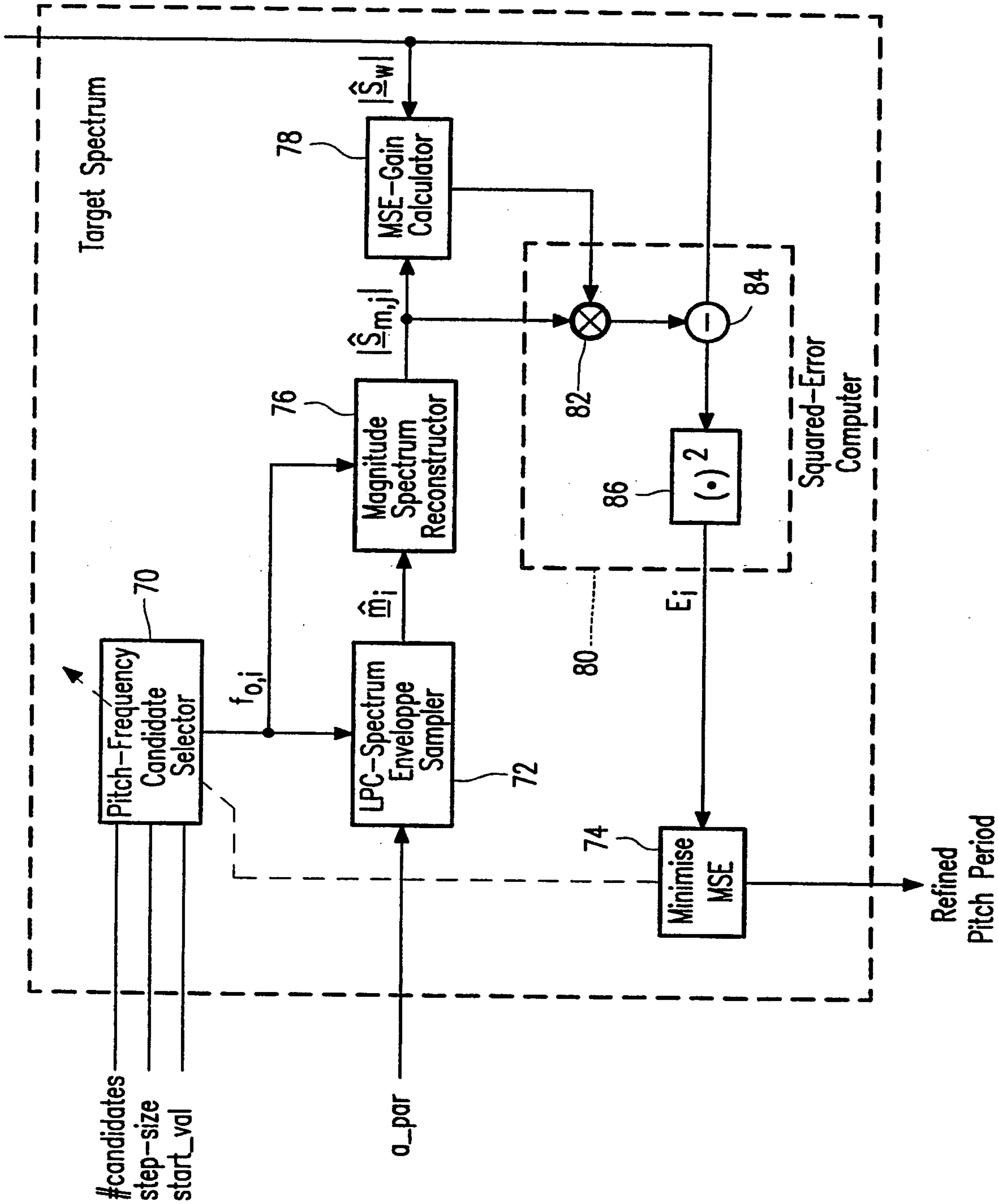

30
Time-Domain Windowing
50
Autocorrelation Function Computer
58
Lag Windowing
60
Schur Recursion
62
Reflection Coefficients Quantizer
64
Reflection Coeffs to a-parameters
66
52
Reflection Coefficients Reconstructor
54
Reflection Coeffs to a-parameters
56
s[n]
V/Uv Decision
$\underline{S}_w$
$\underline{R}_{ss}$
$\hat{\underline{R}}_{ss}$
$\underline{k}$
$\underline{C}$
$\hat{\underline{k}}$
a-parameters
LPC Codes
a-parameters (quantised)

FIG. 4

FIG. 5


Target Spectrum
$|\hat{\underline{S}}_w|$
78
MSE–Gain Calculator
$|\hat{\underline{S}}_{m,j}|$
84
76
Magnitude Spectrum Reconstructor
82
Squared–Error Computer
86
$(\cdot)^2$
70
$\hat{\underline{m}}_i$
80
E_i
Pitch–Frequency Candidate Selector
$f_{o,i}$
LPC–Spectrum Envelope Sampler
72
74
Minimise MSE
Refined Pitch Period
#candidates
step–size
start_val
a_par

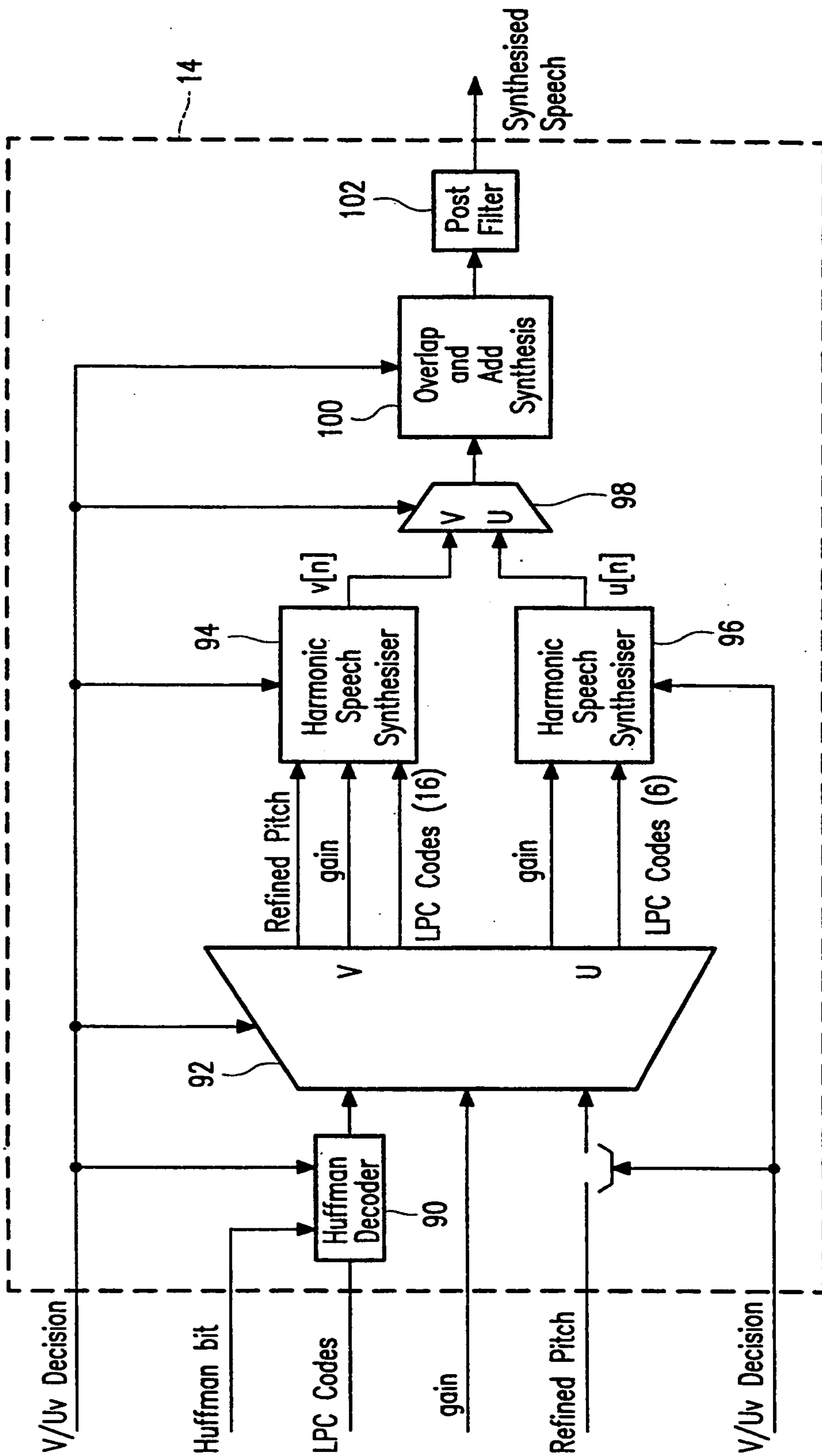

14
Synthesised Speech
102
Post Filter
100
Overlap and Add Synthesis
98
V
U
v[n]
u[n]
94
Harmonic Speech Synthesiser
96
Harmonic Speech Synthesiser
Refined Pitch
gain
LPC Codes (16)
gain
LPC Codes (6)
92
V
U
90
Huffman Decoder
V/Uv Decision
Huffman bit
LPC Codes
gain
Refined Pitch
V/Uv Decision

FIG. 7

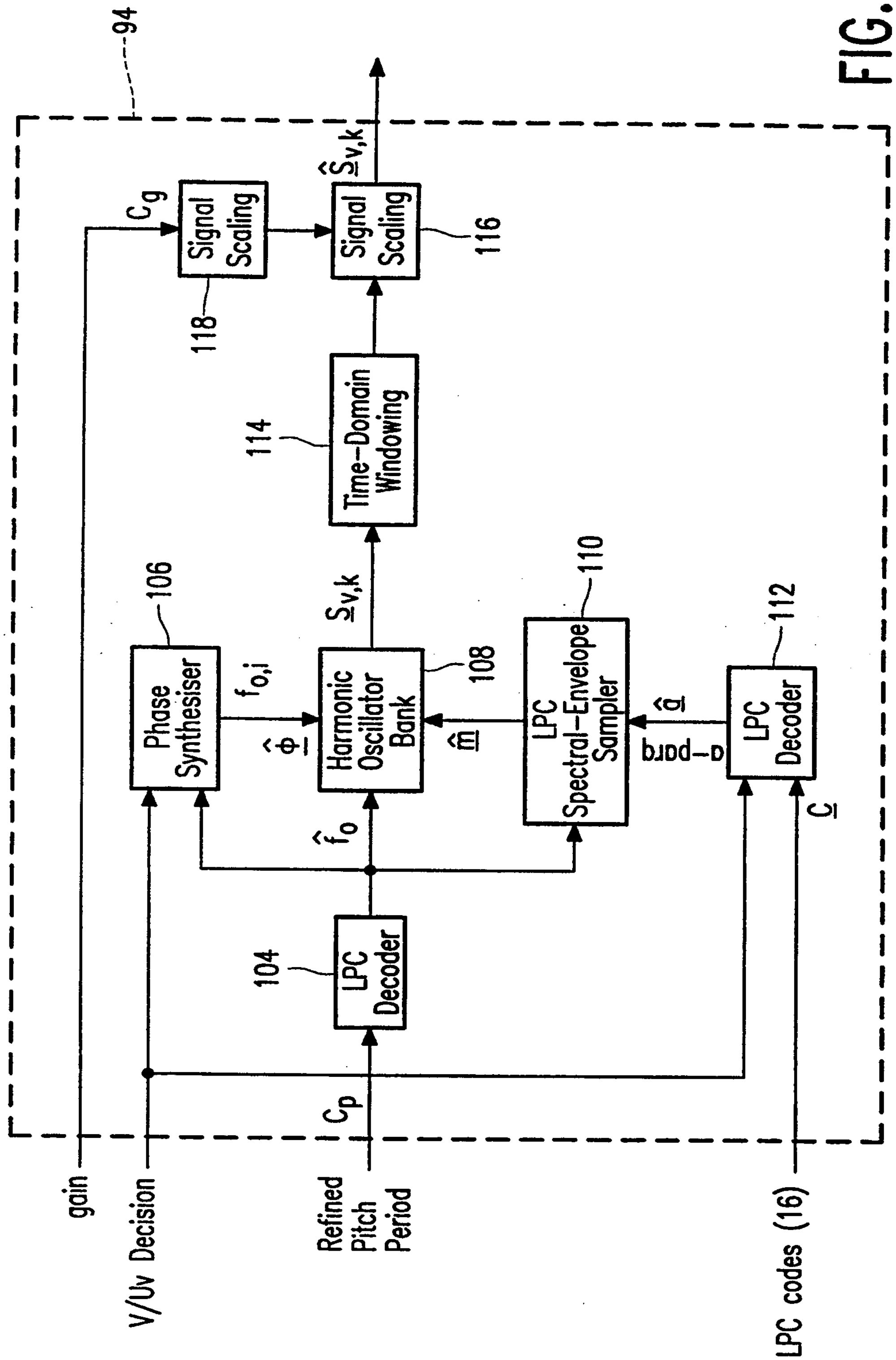

94
gain
V/Uv Decision
Refined Pitch Period
LPC codes (16)
Cp
Cg
C
104
LPC Decoder
106
Phase Synthesiser
fo,i
108
Harmonic Oscillator Bank
110
LPC Spectral-Envelope Sampler
112
LPC Decoder
a-para
114
Time-Domain Windowing
118
Signal Scaling
116
Signal Scaling
Sv,k

FIG. 8

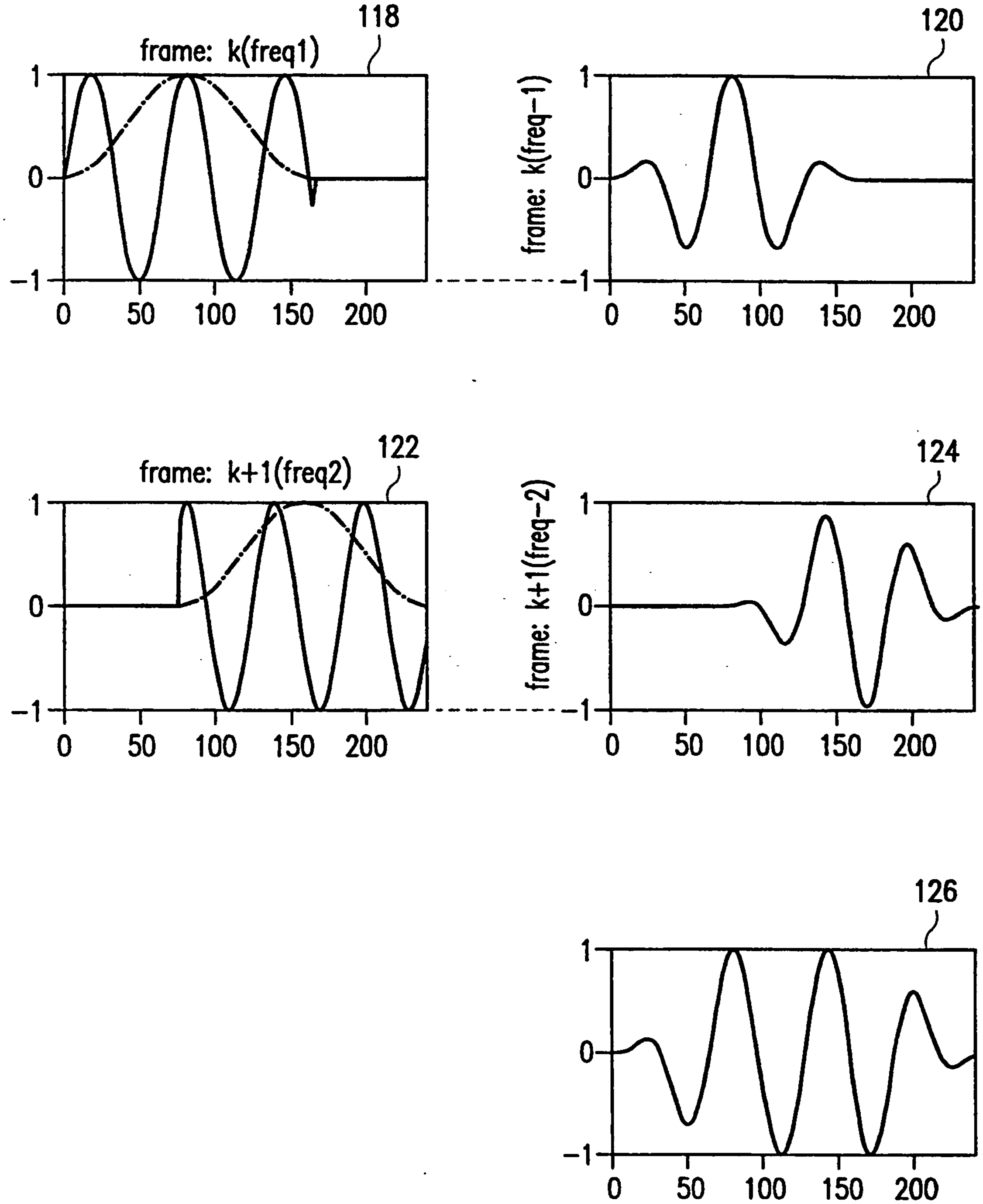

118
frame: k(freq1)
120
frame: k(freq−1)
122
frame: k+1(freq2)
124
frame: k+1(freq−2)
126
1
0
−1
0
50
100
150
200

FIG. 9

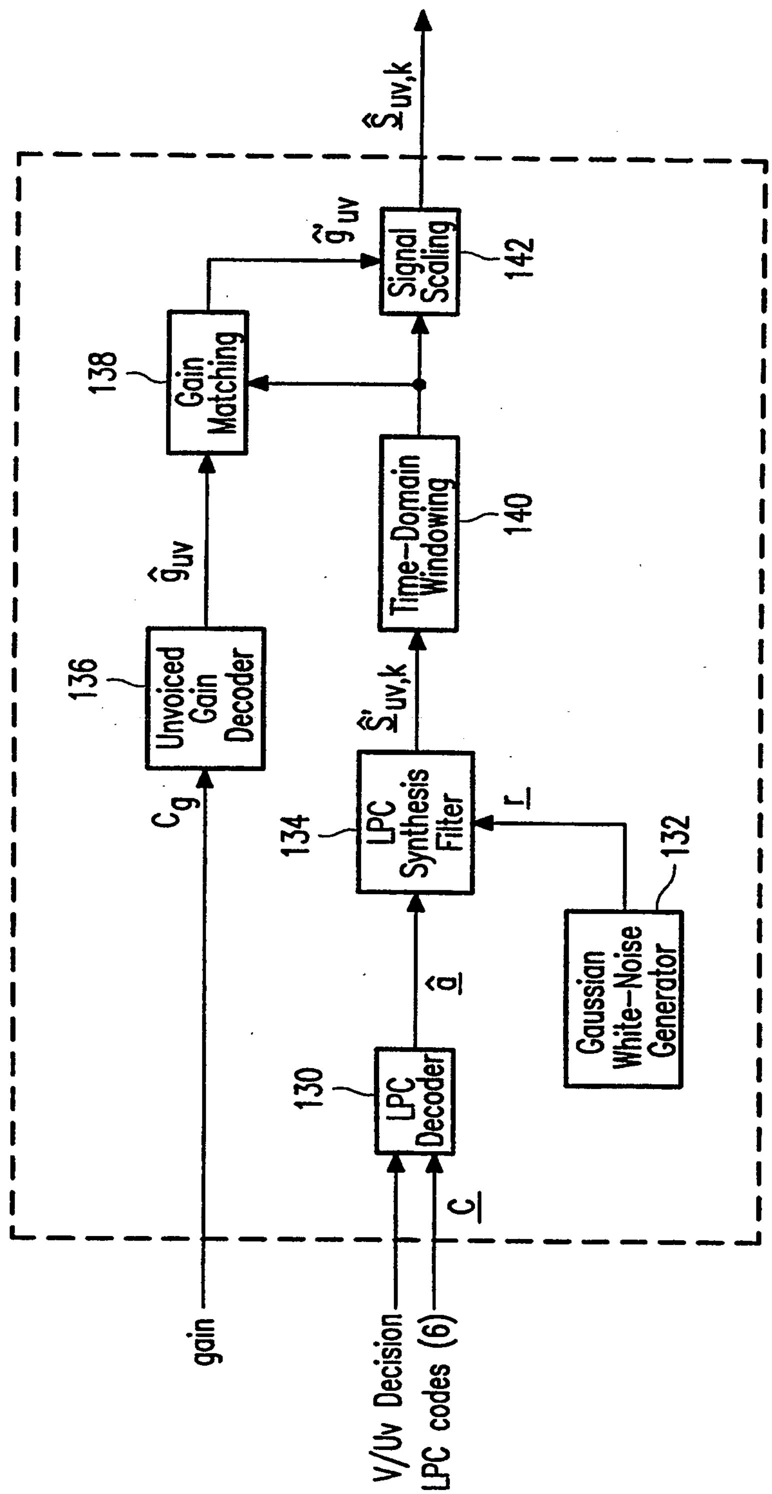

gain
V/Uv Decision
LPC codes (6)
$\underline{C}$
C_g
130
LPC Decoder
$\hat{\underline{a}}$
134
LPC Synthesis Filter
$\underline{r}$
132
Gaussian White-Noise Generator
$\hat{\underline{S}}'_{uv,k}$
136
Unvoiced Gain Decoder
$\hat{g}_{uv}$
140
Time-Domain Windowing
138
Gain Matching
$\tilde{g}_{uv}$
142
Signal Scaling
$\hat{\underline{S}}_{uv,k}$

FIG. 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB 98/00871

A. CLASSIFICATION OF SUBJECT MATTER

IPC6: G10L 3/02, G10L 9/14
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC6: G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0260053 A1 (AMERICAN TELEPHONE AND TELEGRAPH COMPANY), 16 March 1988 (16.03.88), see"solution" --	1-8
X	US 4924508 A (HUBERT CREPY ET AL), 8 May 1990 (08.05.90), column 4, line 52 - column 7, line 55, figure 8, see"summary of the invention" --	1,3,5,6,8
X,P	EP 0837453 A2 (SONY CORPORATION), 22 April 1998 (22.04.98), page 2, line 1 - line 57, abstract --	1-8
A,P	US 5666464 A (MASAHIRO SERIZAWA), 9 Sept 1997 (09.09.97), see"summary of the invention" --	1-8

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" erlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 December 1998	22 -12- 1998
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer Peder Gjervaldsaeter Telephone No. +46 8 782 25 00

INTERNATIONAL SEARCH REPORT

Information on patent family members

01/12/98

International application No.
PCT/IB 98/00871

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0260053 A1	16/03/88	AU 580218 B	05/01/89
		AU 7825487 A	24/03/88
		CA 1307345 A	08/09/92
		DE 3789476 D,T	15/09/94
		JP 2114630 C	06/12/96
		JP 8033754 B	29/03/96
		JP 63070900 A	31/03/88
		KR 9602388 B	16/02/96
		US 4797926 A	10/01/89
US 4924508 A	08/05/90	EP 0280827 A,B	07/09/88
		SE 0280827 T3	
		JP 2505015 B	05/06/96
		JP 63223799 A	19/09/88
EP 0837453 A2	22/04/98	JP 10124094 A	15/05/98
US 5666464 A	09/09/97	CA 2130877 A	27/02/95
		FR 2709367 A,B	03/03/95
		JP 2658816 B	30/09/97
		JP 7064600 A	10/03/95

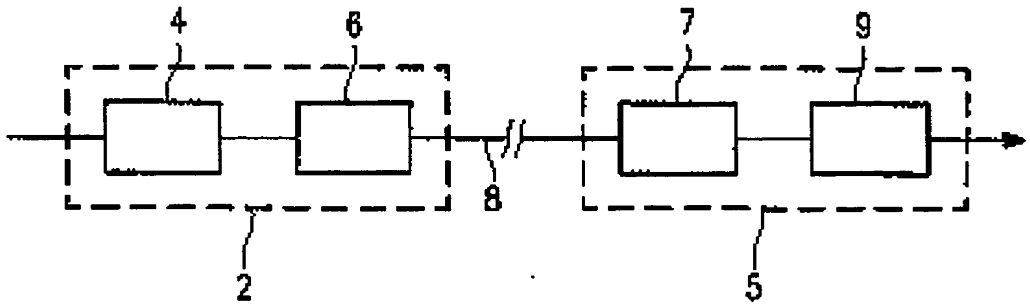

4
6
7
9
8
2
5

FIG. 1

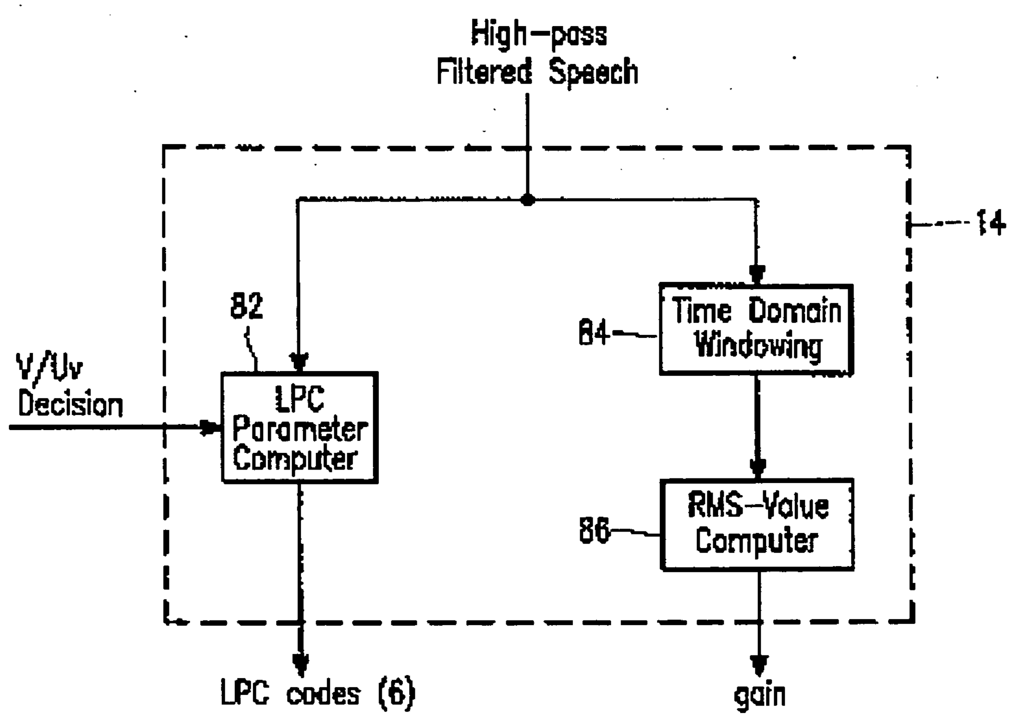

High-pass
Filtered Speech
14
82
84
Time Domain
Windowing
V/Uv
Decision
LPC
Parameter
Computer
86
RMS-Value
Computer
LPC codes (6)
gain

FIG. 6

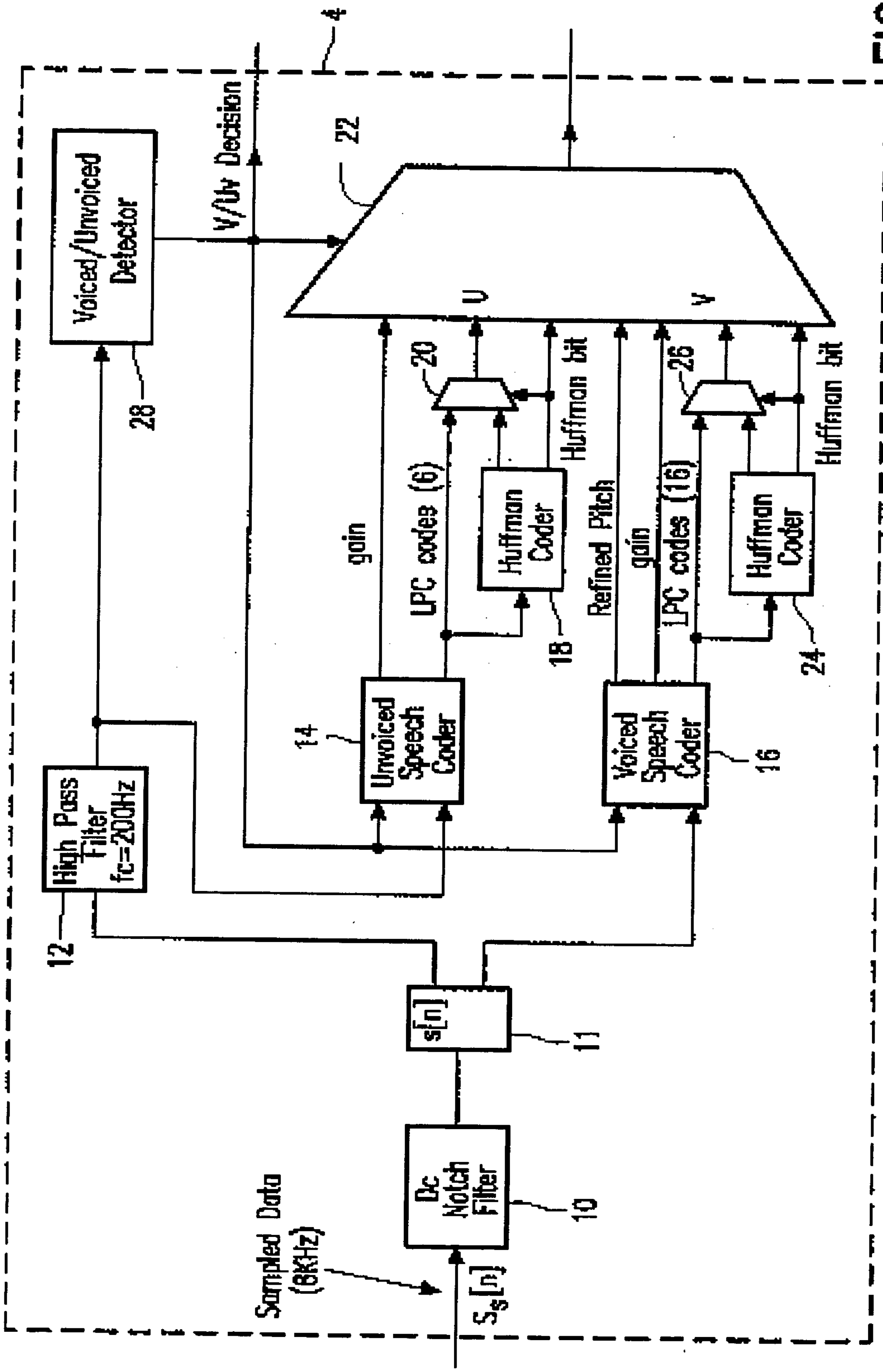

4
Voiced/Unvoiced Detector
28
V/Uv Decision
22
U
V
20
26
gain
LPC codes (6)
Huffman Coder
Huffman bit
18
Refined Pitch
gain
LPC codes (16)
Huffman Coder
Huffman bit
24
High Pass Filter fc=200Hz
12
14
Unvoiced Speech Coder
Voiced Speech Coder
16
s[n]
11
Dc Notch Filter
10
Sampled Data (8KHz)
Ss[n]

FIG. 2

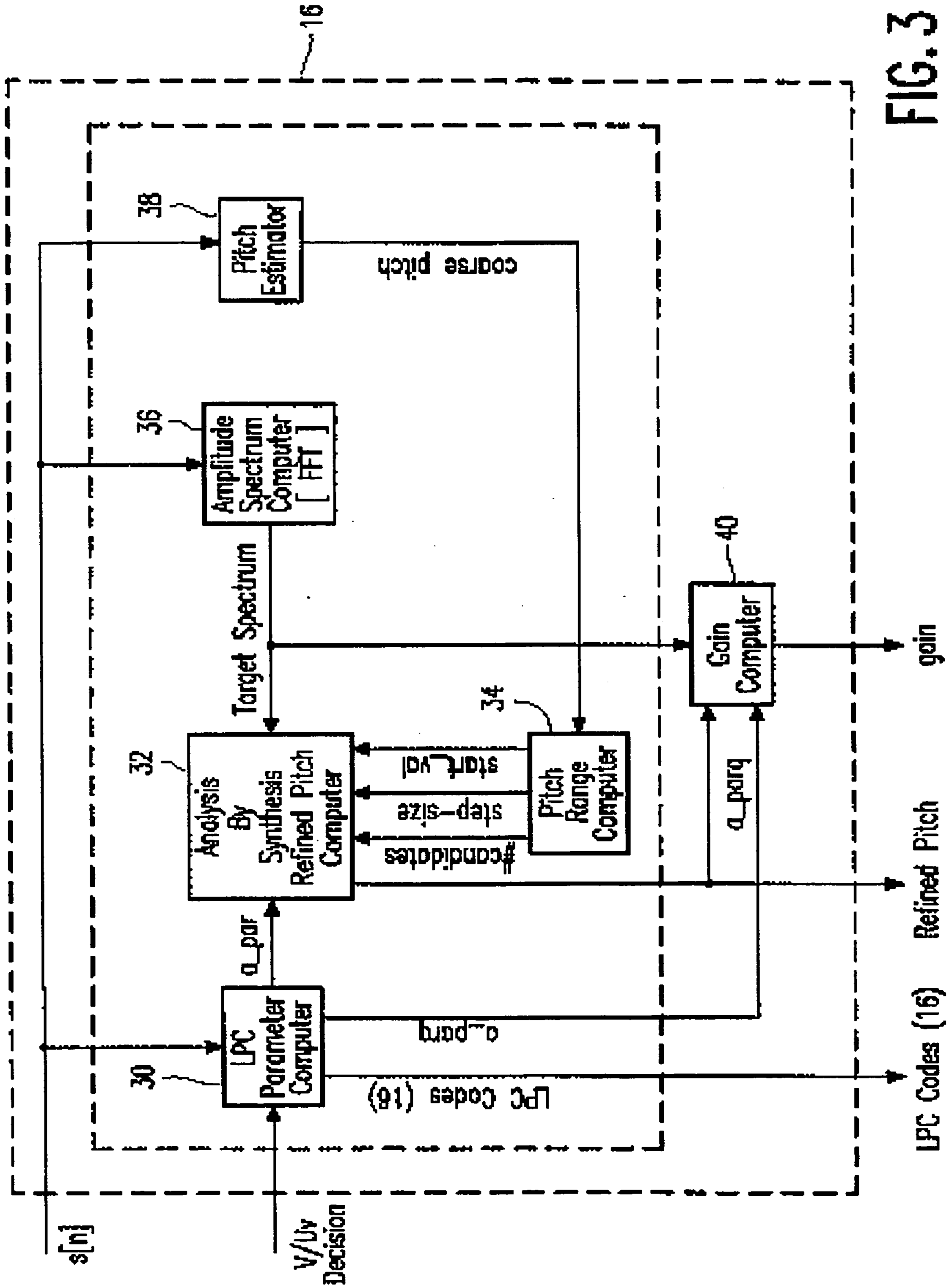

16
38
Pitch Estimator
coarse pitch
36
Amplitude Spectrum Computer [FFT]
Target Spectrum
40
Gain Computer
gain
34
Pitch Range Computer
start_val
step-size
#candidates
32
Analysis By Synthesis Refined Pitch Computer
a_par
a_parq
Refined Pitch
30
LPC Parameter Computer
LPC Codes (16)
s[n]
V/Uv Decision

FIG. 3

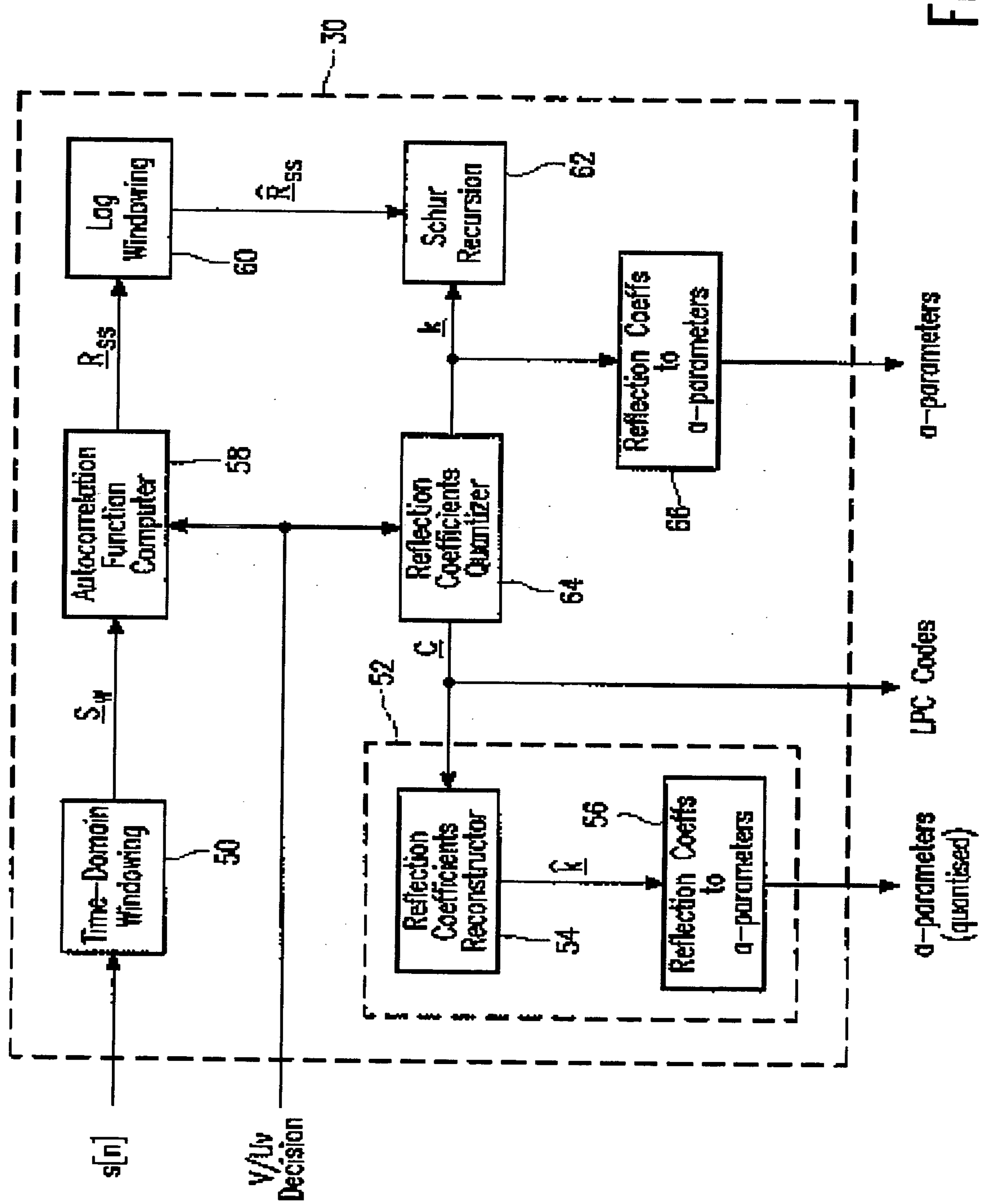


FIG. 4

BNSDOCID: <WO___9903095A1TI_>

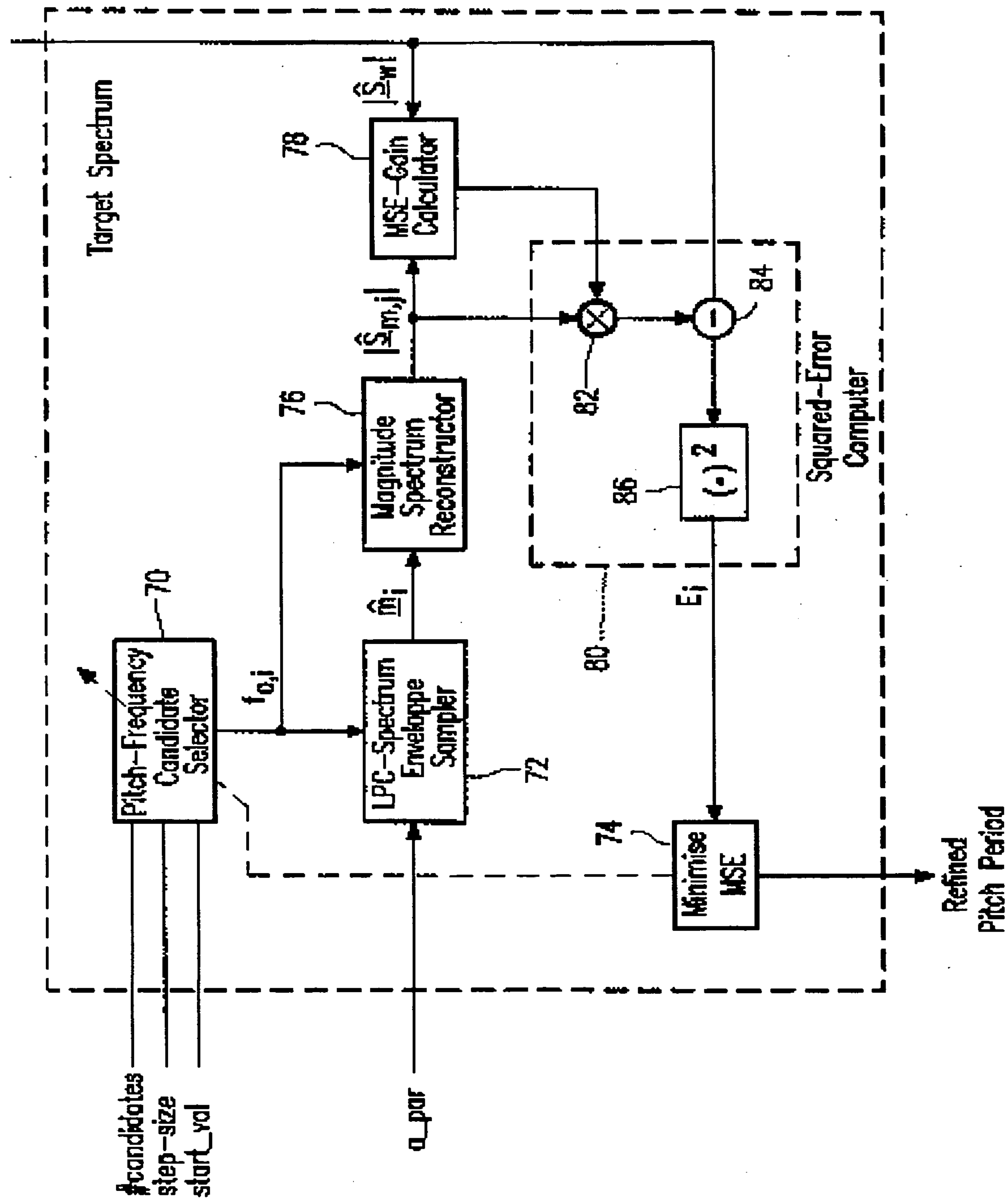

Target Spectrum
Pitch-Frequency Candidate Selector
70
#candidates
step-size
start_val
$f_{o,i}$
LPC-Spectrum Envelope Sampler
72
a_par
$\hat{m}_i$
Magnitude Spectrum Reconstructor
76
$|\hat{S}_{m,i}|$
MSE-Gain Calculator
78
$|\hat{S}_w|$
82
84
$(\cdot)^2$
86
Squared-Error Computer
80
E_i
Minimise MSE
74
Refined Pitch Period

FIG. 5

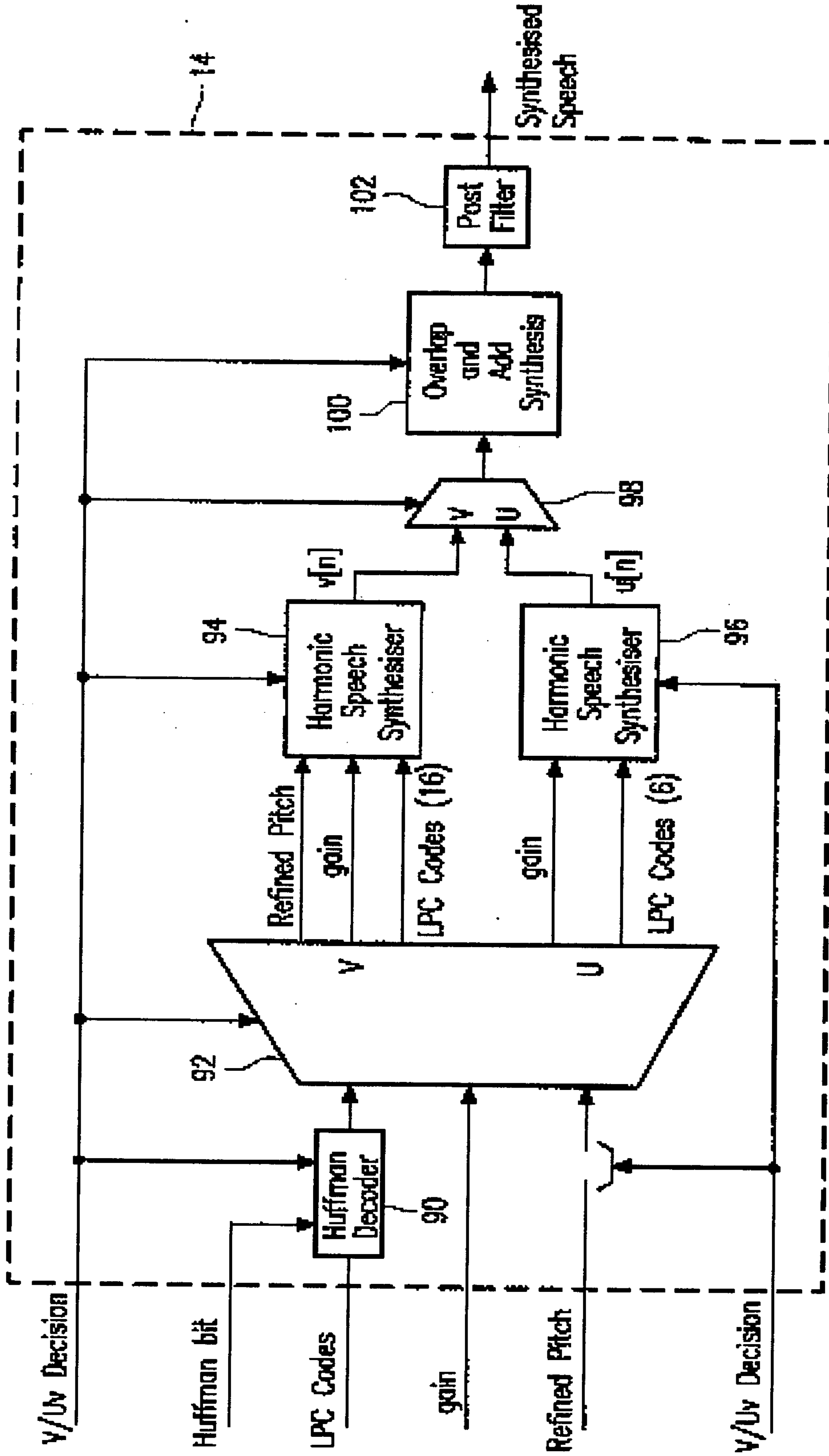

14
Synthesised Speech
102
Post Filter
100
Overlap and Add Synthesis
98
V
U
v[n]
u[n]
94
Harmonic Speech Synthesiser
96
Harmonic Speech Synthesiser
Refined Pitch
gain
LPC Codes (16)
gain
LPC Codes (6)
V
U
92
Huffman Decoder
90
V/Uv Decision
Huffman bit
LPC Codes
gain
Refined Pitch
V/Uv Decision

FIG. 7

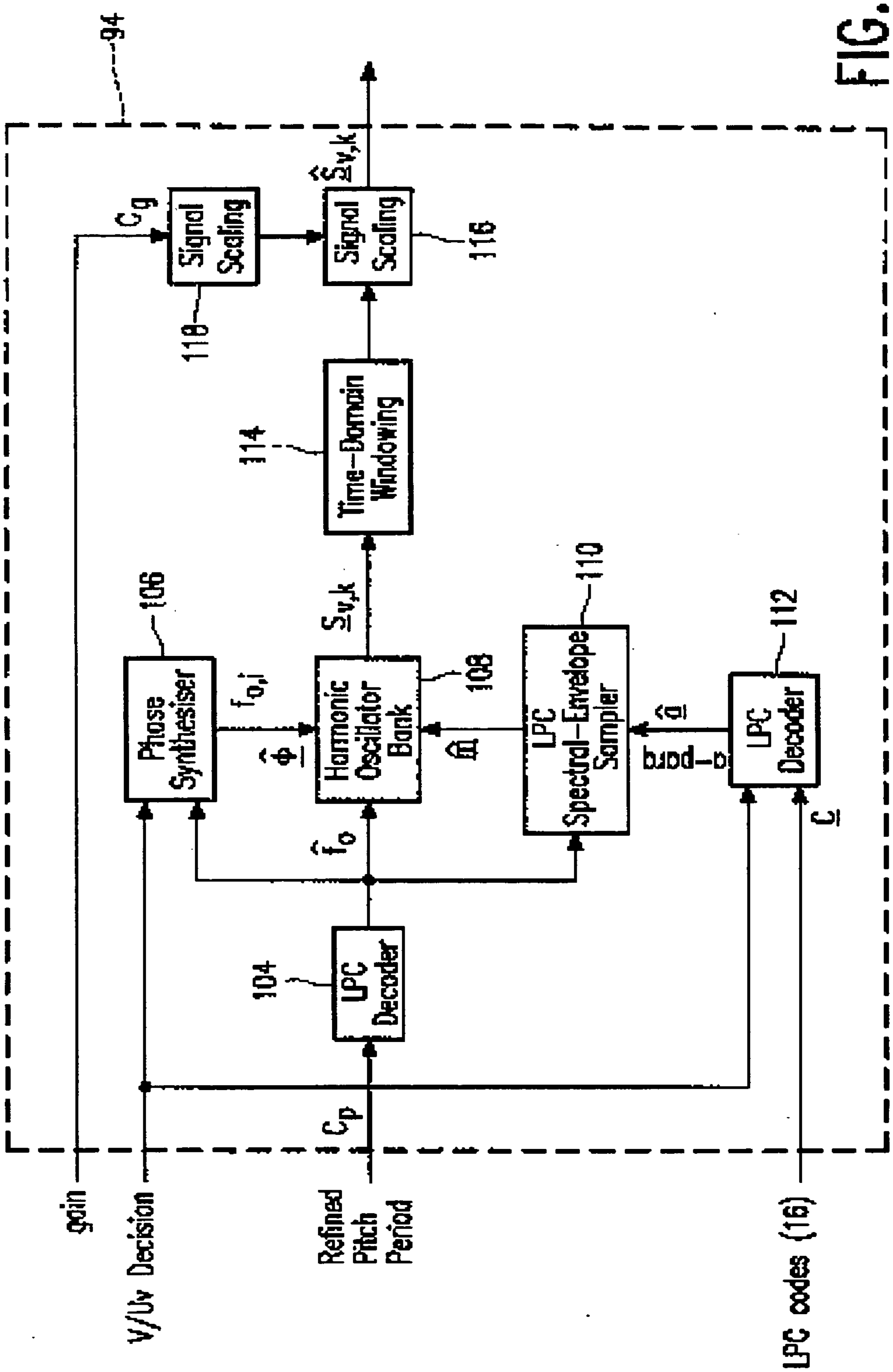
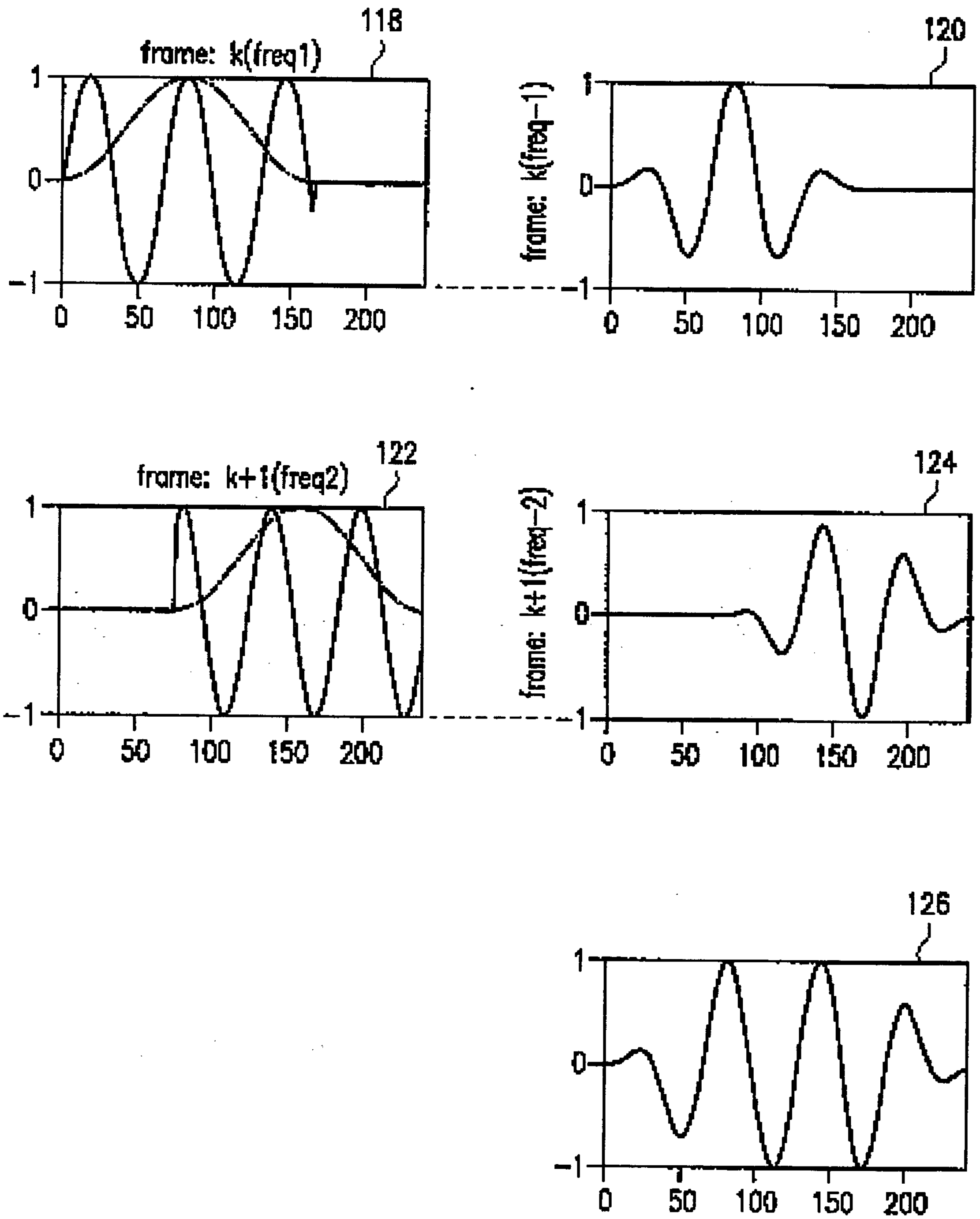

94
Signal Scaling
118
Signal Scaling
116
C_g
$\hat{s}_{v,k}$
Time-Domain Windowing
114
Phase Synthesiser
106
$f_{o,i}$
$\underline{s}_{v,k}$
Harmonic Oscillator Bank
108
$\hat{\underline{\phi}}$
$\hat{\underline{m}}$
LPC Spectral-Envelope Sampler
110
$\hat{\underline{a}}$
LPC Decoder
112
$\underline{c}$
$\hat{f}_o$
LPC Decoder
104
C_p
gain
V/Uv Decision
Refined Pitch Period
LPC codes (16)

FIG. 8



118
frame: k(freq1)
120
frame: k(freq-1)
122
frame: k+1(freq2)
124
frame: k+1(freq-2)
126
1
0
-1
0
50
100
150
200

FIG. 9

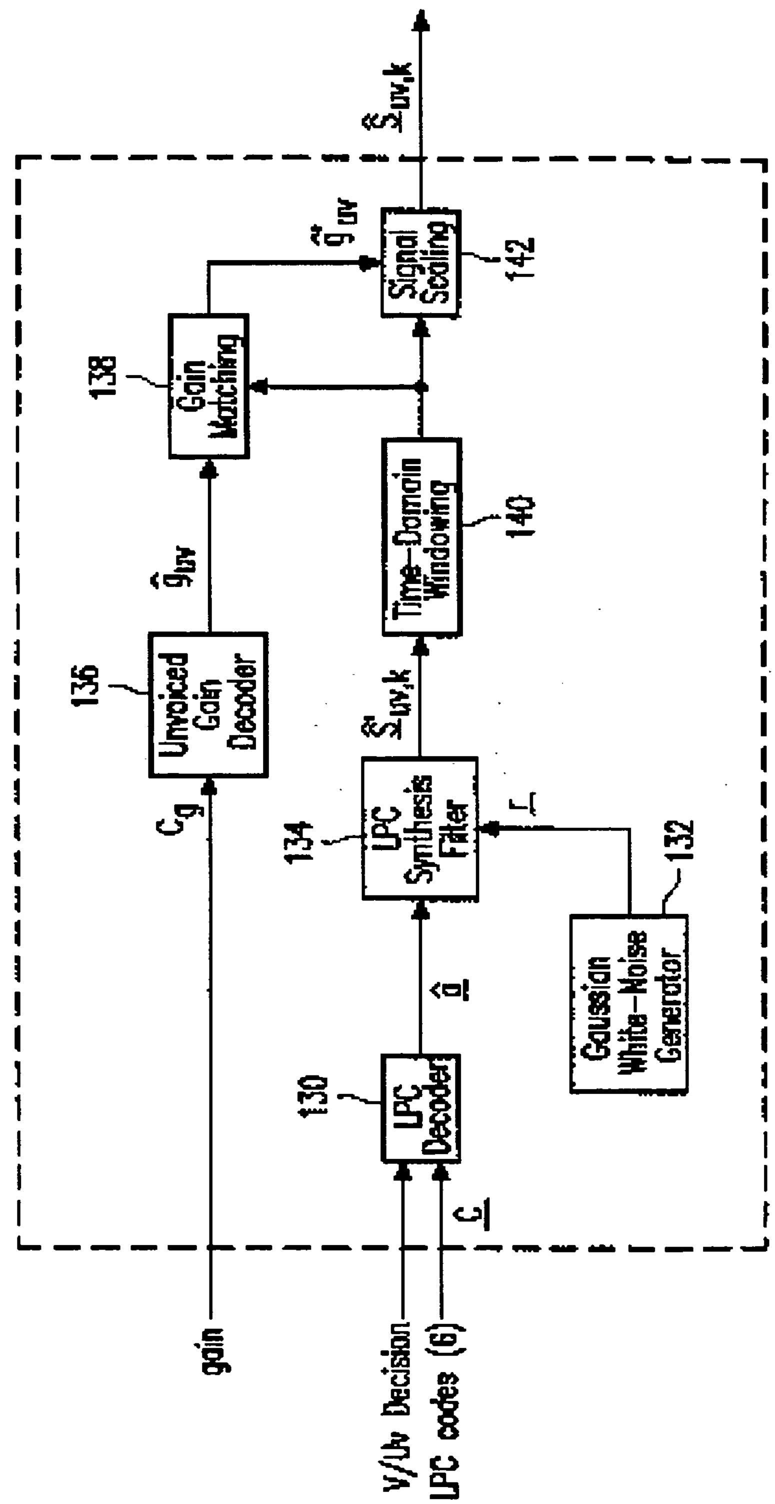

gain
V/UV Decision
LPC codes (6)
$\underline{c}$
C_g
Unvoiced Gain Decoder
136
$\hat{g}_{uv}$
Gain Matching
138
$\hat{\hat{g}}_{uv}$
LPC Decoder
130
$\hat{\underline{a}}$
LPC Synthesis Filter
134
$\underline{r}$
Gaussian White-Noise Generator
132
$\hat{\underline{s}}_{uv,k}$
Time-Domain Windowing
140
Signal Scaling
142
$\hat{\hat{\underline{s}}}_{uv,k}$

FIG. 10

This Page Blank (uspto)